UNIVERSITY OF CAPE TOWN

New Applications of Statistics in Astronomy and Cosmology

by

Michelle Lochner (née Knights)

Thesis presented for the degree of
Doctor of Philosophy

in the
Department of Mathematics and Applied Mathematics
University of Cape Town

September 2014

Declaration of Authorship

I, Michelle Lochner, know the meaning of plagiarism and declare that all of the work in the thesis, save for that which is properly acknowledged, is my own.

Signed:

Date:

"Per ardua ad astra."

Through adversity to the stars

UNIVERSITY OF CAPE TOWN

Abstract

Doctor of Philosophy

New Applications of Statistics in Astronomy and Cosmology

by Michelle Lochner (née Knights)

Over the last few decades, astronomy and cosmology have become data-driven fields. The parallel increase in computational power has naturally lead to the adoption of more sophisticated statistical techniques for data analysis in these fields, and in particular, Bayesian methods. As the next generation of instruments comes online, this trend should be continued since previously ignored effects must be considered rigorously in order to avoid biases and incorrect scientific conclusions being drawn from the ever-improving data.

In the context of supernova cosmology, an example of this is the challenge from contamination as supernova datasets will become too large to spectroscopically confirm the types of all objects. The technique known as BEAMS (Bayesian Estimation Applied to Multiple Species) handles this contamination with a fully Bayesian mixture model approach, which allows unbiased estimates of the cosmological parameters. Here, we extend the original BEAMS formalism to deal with correlated systematics in supernovae data, which we test extensively on thousands of simulated datasets using numerical marginalization and Markov Chain Monte Carlo (MCMC) sampling over the unknown type of the supernova, showing that it recovers unbiased cosmological parameters with good coverage.

We then apply Bayesian statistics to the field of radio interferometry. This is particularly relevant in light of the SKA telescope, where the data will be of such high quantity and quality that current techniques will not be adequate to fully exploit it. We show that the current approach to deconvolution of radio interferometric data is susceptible to biases induced by ignored and unknown instrumental effects such as pointing errors, which in general are correlated with the science parameters. We develop an alternative approach - Bayesian Inference for Radio Observations (BIRO) - which is able to determine the joint posterior for all scientific and instrumental parameters. We test BIRO on several simulated datasets and show that it is superior to the standard CLEAN and source extraction algorithms. BIRO fits all parameters simultaneously while providing unbiased estimates - and errors - for the noise, beam width, pointing errors and the fluxes and shapes of the sources.

Contents

List of Figures

List of Tables

Abbreviations

SKA	Square Kilometre Array
FLRW	Friedmann-Lemaître-Robertson-Walker
ΛCDM	Cosmological constant and Cold Dark Matter
CMB	Cosmic Microwave Background
BAO	Baryon Acoustic Oscillations
MCMC	Markov Chain Monte Carlo
BEAMS	Bayesian Estimation Applied to Multiple Species
SN	supernova
LSST	Large Synoptic Survey Telescope
MSE	Mean Squared Error
RA	Right Ascension
DEC	Declination
BIRO	Bayesian Inference for Radio Observations
RIME	Radio Interferometry Measurement Equation
WSRT	Westerbork Synthesis Radio Telescope
QMC	Quality Monitoring Committee
PyBDSM	Python Blob Detection and Source Measurement software

Physical Constants

Speed of Light	c	=	$2.997\ 924\ 58 \times 10^{8}$ m s^{-1}
Gravitational Constant	G	=	$6.673\ 84 \times 10^{-11}$ m^3 kg^{-1} s^{-2}
Planck constant	h	=	$6.626\ 069\ 57 \times 10^{-34}$ m^2 kg s^{-1}
Boltzmann constant	k_B	=	1.380 648 8 m^2 kg s^{-2} K^{-1}
Solar mass	$M_\odot$	=	1.989×10^{30} kg

Symbols

ν	frequency	s^{-1}
λ	wavelength	m^{-1}
$P(\theta\|D)$	posterior (probability of θ *given* D)	Defined on Page 56
$\mathcal{L}(\theta)$	likelihood	Defined on Page 56
$\boldsymbol{J}$	Jones matrix	Defined on Page 99
V_{pq}	visibility matrix	Defined on Page 100
B	brightness matrix	Defined on Page 101

Acknowledgements

I would like to thank:

My husband,

> who always stood by me, truly listened when I complained about having a multi-modal posterior, and when all else failed, brought me endless cups of tea,

my Mom,

> who supported my Physics career so much, she now phones me on Sundays to ask about quantum physics and dynamical dark energy, and who also proof-read my entire thesis... twice,

my Dad, Grandpa and sisters,

> who first introduced me to the stars and who have never wavered in supporting me,

my Grandma,

> who only believed astronomy was a real career when she saw an astronomer in a movie,

Lise, Eli, Ermias, Rene, Nadeem, Kai, Navin, Mark, Iggy,

> my colleagues, office mates and friends, for reading my thesis and giving comments, for providing the daily distractions and discussions that make being a student fun, and for making AIMS the great place it is,

Martin Kunz,

> for funding a research visit to Geneva, for comments on my thesis and for being a guru of Bayesian statistics,

Filipe Abdallah and Oleg Smirnov

> for giving me the occasional crash course on radio astronomy, for helping me find my way through the forest that is MeqTrees, for giving up way too many hours of their time to install MeqTrees on the UCT cluster,

and of course, my supervisor Bruce Bassett,

> who taught me when I was hungry to learn, who pushed me when I was afraid of moving forward and who was kind to me when I needed it most. Everything that I am or will become as a researcher, is thanks to what you have given me.

Formal Acknowledgements

Substantial sections of the BEAMS Chapter are drawn from the paper [1], of which I am first author. My coauthors Bruce Bassett, Melvin Varughese, Renée Hlozek, Martin Kunz, Mat Smith and James Newling all contributed to the development and guidance of this project and I am grateful for their support and comments on the original paper.

The BIRO chapter forms part of a project in collaboration with Oleg Smirnov, Bruce Bassett, Jon Zwart, Martin Kunz, Nadeem Oozeer and Iniyan Natarajan. The work done is part of a draft paper, of which I am also first author, which we hope to publish soon. I thank my colleagues for their contributions in the design, development and implementation of this project.

I acknowledge funding from the NRF/SKA and from the University of Cape Town and support and resources from the African Institute for Mathematical Sciences. Part of the numerical calculations for this thesis were performed on the Andromeda cluster of the University of Geneva. Other computations were performed using facilities provided by the University of Cape Town's ICTS High Performance Computing team: http://hpc.uct.ac.za.

Preface

The versatility and applicability of Bayesian statistics on a multitude of problems is both an advantage and disadvantage. The advantage I have enjoyed during my PhD in Cape Town is the ability to work effectively on a variety of interesting fields. The disadvantage is trying to capture that large scope sufficiently and sensibly into a single thesis. The subsequent chapters may thus seem disconnected at first glance, but they do all link to the main body of research, which is applying Bayesian statistics to deal with systematic errors that may cause problems for current and future astronomical observations.

Chapters 1, 2 and 3 are introductory chapters, covering the theory and background required for Chapters 4 and 5: the research chapters. The first part of Chapter 1 is a fairly general introduction to cosmology to put the rest of the thesis into context and to allow the framework in which to develop the scientific background of later sections. This chapter then moves on to discuss supernova cosmology, which is used in Chapter 4 where BEAMS is applied, and the epoch of reionization and large scale structure, which is indirectly relevant to BIRO (Chapter 5) because these are topics of interest to the SKA[1]. The last section of Chapter 1 is a brief overview of sources of radio frequency observations, which can be analysed using BIRO. Chapter 2 focuses on the nitty gritty of radio astronomy observations, required to develop the BIRO technique in Chapter 5. Of course, the overview of Bayesian statistics, Chapter 3, is used extensively by both research chapters.

[1]Square Kilometre Array, www.skatelescope.org

Chapter 1

An Introduction to Astronomy and Cosmology

1.1 Introduction

> *Man must rise above the Earth to the top of the atmosphere and beyond for only thus will he fully understand the world in which he lives.*
>
> - Socrates

Cosmology is the study of the Universe: how it began, how it is evolving and how (and if) it will end. It is a wide, far-reaching field which has captured the imagination of scientists and philosophers alike for centuries. In the last two decades, cosmology has moved from a qualitative, purely theoretical science, to a quantitative, data-driven field, as our experiments and telescopes become larger and more ambitious.

The study of cosmology stretches from the smallest scales, such as the particle physics of the early Universe, to the largest, such as the scales of galaxy superclusters and the question of the size of the Universe itself. Central to the standard model of cosmology is the cosmological principle which is the idea that viewed on a sufficiently large scale, the Universe has the same properties for all observers [2, 3].

The cosmological principle is a more specific form of the Copernican principle, which is the idea that Earth does not occupy a special place in the Universe. While this idea may be widely

accepted in modern theories, this was not so at the time of Copernicus and later Bruno and Galileo, who fought the view that the Earth was at the centre of the Universe [4].

This principle is valid on scales of around 100 Mpc, where the Universe appears to be quite smooth (and hence probably looks roughly the same to any observer in the Universe). As well as an important philosophical idea, it is also a central assumption to most models of the Universe.

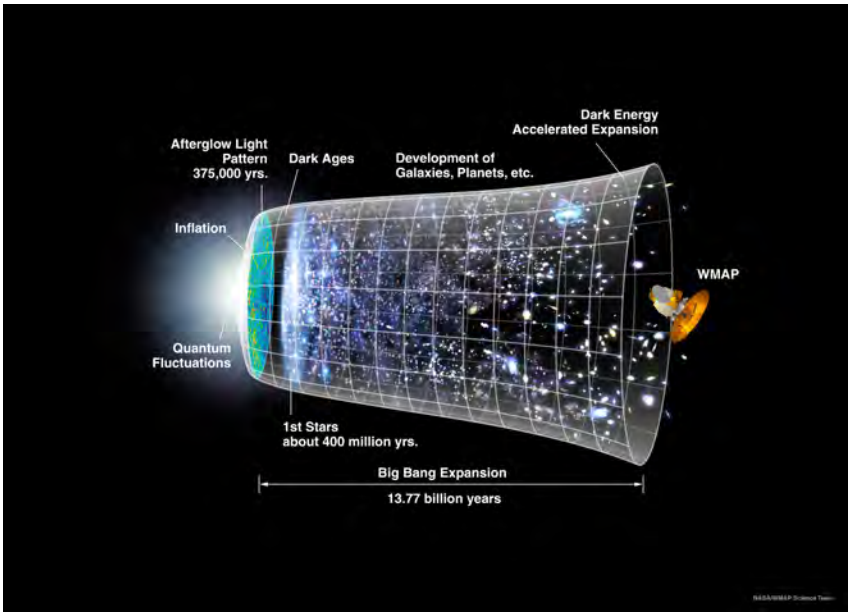


FIGURE 1.1: A timeline of the standard model of the evolution of the Universe. The size of the observable Universe is represented by the vertical extent of the grid on the image. Credit: NASA / WMAP Science Team.

From the foundations of the cosmological principle, scientists have spent decades piecing together the story of the Universe. Figure 1.1 is a overview of our understanding of cosmic history: from the Big Bang, through inflation, structure formation and accelerated expansion. It took almost the whole 20th century to build this picture.

The story began with Einstein's development of General Relativity [**?**], which led to our modern understanding of gravity. The Big Bang theory of the beginning of the Universe (that the Universe had not always existed but instead started in a hot, dense state), developed by George Lemaître and George Gamow [5], was initially dismissed by the scientific community. This changed with Vesto Slipher and Edwin Hubble's findings [6, 7], that all galaxies appeared to be moving away from us, was a powerful piece of evidence towards an expanding Universe and hence the Big Bang theory. Thanks mostly to the accidental discovery of the cosmic microwave background (CMB) radiation [3] (the afterglow radiation from the Big Bang) in 1965, the Big Bang is now accepted by scientific consensus.

In addition to the precise measurements of the CMB, the study of structure formation and matter in the Universe have also revolutionised the field of observational cosmology [8]. This field, by the end of the 20th century, was left in an odd state. Observations suggest that all known matter, including stars, planets and dust, makes up a mere 4% of the energy content of the Universe. About 23% is in the form of dark matter, an unfamiliar form of matter which interacts only gravitationally with normal matter. The remaining 73% is dark energy, hypothesised to be an unknown material which causes the late-time acceleration of the expansion of the Universe (the upward turn in Figure 1.1). While little is known about dark matter, even less is known of dark energy and theories abound to describe both of these or provide alternatives (see Section 1.5.3) [8].

In this chapter, we cover some of the fundamental theory of cosmology, starting from General Relativity and moving to describing the geometry and content of the Universe. We then focus on some observational cosmology topics relevant to the research in this thesis. Supernova cosmology is relevant to the research of Chapter 4. The remaining sections, the epoch of reionization and large scale structure, are indirectly relevant to Chapter 5 in that these are essential fields of study for the SKA[1] and are thus useful science to consider in the development of the technique BIRO from Chapter 5.

1.2 General Relativity

Einstein's General Theory of Relativity [**?**] is conceptually simple, but mathematically sophisticated. It is the theory of how spacetime, curvature and gravity are related. Einstein's insight was that while other fundamental forces are represented by fields defined *on* spacetime, gravity is a property *of* spacetime itself [9]. A detailed study of General Relativity in this thesis is impossible so we will focus on aspects central to cosmology, with some necessary background material. The reader is referred to standard textbooks for more detail (e.g. [9, 10]). The material for this section, unless otherwise stated, is drawn from [9].

We start with Einstein's equation (Eq.(1.1)): a deceptively simple equation which encodes the essentials of the theory. Einstein's equation was not derived from first principles, rather Einstein postulated it using logical argument while trying to construct a theory which supercedes basic

[1] Square Kilometre Array, http://www.skatelescope.org

Newtonian gravity. Since Einstein first proposed his theory in 1916, General Relativity has stood up to every observational test [11].

Einstein's equation (in tensor form) is given by:

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = 8\pi G T_{\mu\nu}, \tag{1.1}$$

where G is the gravitational constant and where we have set the speed of light, c, to one and will continue to do so throughout this section. In layman's terms, Einstein's equation says: "Spacetime tells matter how to move, matter tells spacetime how to curve"[2]. The terms on the left of the equation describes the curvature of spacetime while the term on the right describes matter. We will now briefly explain each term in the Einstein equation.

$R_{\mu\nu}$ is called the Ricci tensor[3] and is derived by contracting the Riemann tensor (Eq.(1.8)). R is the Ricci scalar which is obtained by a further contraction on the indices (Eq.(1.9)). In order to understand where the Ricci tensor and Ricci scalar come from, we need a few new concepts and equations.

The first is the important concept of the metric tensor, written as $g_{\mu\nu}$, which defines the geometry of a manifold (a topological space which may be curved, but locally resembles Euclidean space) [14]. The metric is essential in defining the spacetime interval between events (also called a line element):

$$ds^2 = g_{\mu\nu}dx^\mu dx^\nu, \tag{1.2}$$

where repeated indices indicates a sum over those indices. This is an invariant quantity: even if we change the coordinate system, the spacetime interval will not change.

If we consider the usual three spatial dimensions and one time dimension, then x^μ can be written as a vector: $x^\mu = \{t, x, y, z\}$. If we further only consider Minkowski space (the equivalent of Euclidean space where time is included as a fourth dimension), then the metric, using the special

[2] Quoted from John Archibald Wheeler in his autobiography [12].

[3] We assume the reader is familiar with the concept of a tensor. See [9, 13] or any other textbook on general relativity for an introduction to tensor calculus.

symbol $\eta_{\mu\nu}$, is written as[4]:

$$\eta_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{1.3}$$

such that, in Minkowski space, the spacetime interval is:

$$ds^2 = -(dt)^2 + (dx)^2 + (dy)^2 + (dz)^2. \tag{1.4}$$

Another essential piece of the Einstein's equation puzzle is the Christoffel symbols [9]. These are determined from the metric by:

$$\Gamma^{\lambda}{}_{\mu\nu} = \tfrac{1}{2} g^{\lambda\sigma}(\partial_\mu g_{\nu\sigma} + \partial_\nu g_{\sigma\mu} - \partial_\sigma g_{\mu\nu}). \tag{1.5}$$

The Christoffel symbols are essential in defining the covariant derivative, ∇, an operator which performs the function of the partial derivative in curved space but in a coordinate independent way. It is defined, on arbitrary vector V^μ, as:

$$\nabla_\mu V^\nu = \partial_\mu V^\nu + \Gamma^{\nu}{}_{\mu\lambda} V^\lambda. \tag{1.6}$$

The Christoffel symbols are also needed in the Riemann tensor, which describes the curvature of the manifold [9]:

$$R^{\rho}{}_{\sigma\mu\nu} = \partial_\mu \Gamma^{\rho}{}_{\nu\sigma} - \partial_\nu \Gamma^{\rho}{}_{\mu\sigma} + \Gamma^{\rho}{}_{\mu\lambda}\Gamma^{\lambda}{}_{\nu\sigma} - \Gamma^{\rho}{}_{\nu\lambda}\Gamma^{\lambda}{}_{\mu\sigma}. \tag{1.7}$$

Finally, the Riemann tensor can be contracted to form the Ricci tensor:

$$R_{\mu\nu} = R^{\lambda}{}_{\mu\lambda\nu}, \tag{1.8}$$

and the Ricci scalar [9]:

$$R = R^{\mu}{}_{\mu} = g^{\mu\nu} R_{\mu\nu}. \tag{1.9}$$

[4]Here we use the sign convention [-1,1,1,1] [9].

These are all the elements which make up the left hand side of Eq.(**??**), together describing the curvature of spacetime.

We now turn our attention to the right hand side of Einstein's equation which describes matter and energy.

The energy-momentum tensor (or the stress-energy tensor), $T_{\mu\nu}$, is defined as the flux of four-momentum p^μ across a surface of constant x^ν [9]. To illustrate the use of this tensor, let us consider a perfect fluid, that is a fluid which can be completely described by two quantities: the rest-frame energy density, ρ, and an isotropic rest-frame pressure, p. The rest-frame form of the energy-momentum tensor is then

$$T_{\mu\nu} = \begin{pmatrix} \rho & 0 & 0 & 0 \\ 0 & p & 0 & 0 \\ 0 & 0 & p & 0 \\ 0 & 0 & 0 & p \end{pmatrix}. \tag{1.10}$$

In general, this is written as [14]:

$$T^{\mu\nu} = (\rho + p)U^\mu U^\nu + p\eta^{\mu\nu}, \tag{1.11}$$

where $U^\mu = \frac{dx^\mu}{dt}$ is the four-velocity of the fluid and $\eta^{\mu\nu}$ is the Minkowski metric (Eq.(1.3)). The reader is also reminded that the inverse of a metric is written with indices raised and subject to the relation $\eta^{\mu\nu}\eta_{\nu\rho} = \delta^\mu{}_\rho$[9]. We are able to model many forms of matter as a perfect fluid in our description of the Universe. We can solve for the evolution of a perfect fluid from its equation of state: $p = p(\rho)$. Dust (as we describe all matter in the Universe) is pressureless ($p = 0$) while an isotropic gas of photons has $p = \frac{1}{3}\rho$ [3]. One last important point about the energy-momentum tensor is that it is conserved: $\nabla_\mu T^{\mu\nu} = 0$. This is by construction in Einstein's equation to ensure conservation of energy and momentum [9].

General relativity is quite different from non-gravitational theories (such as special relativity) in that the actual value of the energy field matters, not just changes in it [9]. This allows the possibility of a vacuum energy, which is the energy density of empty space. The vacuum energy must have a negative pressure, $p_{\text{vac}} = -\rho_{\text{vac}}$, to ensure $T_{\mu\nu}$ from Eq.(1.11) is Lorentz invariant (i.e. the vacuum should be the same in all frames).

When Einstein wrote down his theory, he tried to find a static cosmological model, as it was the consensus at the time that the Universe was unchanging. However, his equations implied that a Universe with uniform matter density distribution cannot be static, unless a finely tuned cosmological constant, Λ, is introduced:

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi G T_{\mu\nu} \tag{1.12}$$

However, this static solution was proven to be unstable, at least in the case of a matter dominated Universe [15]. Einstein reputedly called this addition of the cosmological constant "his biggest blunder" when observations by Vesto Slipher and Edwin Hubble [6, 7] showed that the Universe was expanding and a solution to the original Einstein equation was found which allowed for an expanding Universe. Towards the end of the century, however, the cosmological constant was revived as an excellent description for dark energy (see Section 1.5.3 for more details) [8].

1.3 The Friedmann-Lemaître-Robertson-Walker Universe

One of the most fundamental assumptions in modern cosmology is the cosmological principle, the result of which is that the Universe is isotropic (invariant under rotation) around every point and therefore homogeneous (invariant under spatial translation) [8]. This is a more specific version of the Copernican principle (dating back to Nicolaus Copernicus), which is the idea that we do not live in a special place in the Universe. While there is ample evidence for isotropy from the CMB [3], homogeneity is harder to prove and inhomogeneous models are an active field of research (see for example [16]). As the assumption of homogeneity not only makes calculations much easier, as well as reasonable and philosophically attractive, we will utilise it throughout this work [3, 8].

The metric for a homogeneous and isotropic spacetime is called the Friedmann-Lemaître-Robertson-Walker (FLRW) metric [9] and is given by:

$$ds^2 = -dt^2 + a^2(t)\left[\frac{dr^2}{1-\kappa r^2} + r^2 d\Omega^2\right], \tag{1.13}$$

where

$$d\Omega^2 = d\theta^2 + sin^2(\theta)d\phi^2 \tag{1.14}$$

and $a(t)$ is the dimensionless scale factor. This scale factor relates the distance between two galaxies at some time, t, in the past to the distance at reference time t_0. The curvature parameter, $\kappa = k/R_0^2$, describes the curvature of spacetime, where R_0 is the radius of curvature at $a(t_0) = 1$. The fixed parameter k is $+1$ if the Universe has constant positive curvature, 0 if it is flat and -1 refers to constant negative curvature. See Section 1.5.4 for more details on curvature.

Lastly, the r in Eq.(1.13) is $r = R_0 S_k(\chi)$, where

$$S_k(\chi) = \begin{cases} \sin(\chi) & \text{if } k = +1 \\ \chi & \text{if } k = 0 \\ \sinh(\chi) & \text{if } k = -1 \end{cases} \tag{1.15}$$

and χ is the comoving distance. Thus the FLRW metric describes an expanding, homogeneous and isotropic Universe.

To determine the scale factor, $a(t)$, in Eq.(1.13), we need to use Einstein's equation (Eq.(1.1)) and the perfect fluid form of the energy-momentum tensor (Eq.(1.10)). First, consider the time-time component of the energy-momentum tensor. Since under the assumption of a perfect fluid, the equation of state is $p = w\rho$ (where w is not a function of ρ), conservation of energy implies [9]:

$$\nabla_\mu T_0^\mu = 0. \tag{1.16}$$

Substituting in the FLRW metric leads to:

$$\frac{\dot{\rho}}{\rho} = -3(1+w)\frac{\dot{a}}{a}. \tag{1.17}$$

If we further assume that w is constant, this equation can be integrated to obtain:

$$\rho \propto a^{-3(1+w)}. \tag{1.18}$$

This governs the evolution of a perfect fluid as the scale factor changes.

Now we turn our attention to $a(t)$ itself. Since we have chosen a metric to work with as well as the form of $T_{\mu\nu}$, we can substitute these into Einstein's equation and compute each term of the

tensor equation [9]. We end up with two distinct equations:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{\kappa}{a^2}, \tag{1.19}$$

and

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3p). \tag{1.20}$$

These are known as the Friedmann equations and describe the evolution of the Universe.

From the geodesic equation assuming the FLRW metric (see [17] for a more general derivation), we know that the energy E_1 of a photon emitted at scale factor a_1 is related to the energy E_2 of the photon at scale factor a_2 by [9]:

$$\frac{E_1}{E_2} = \frac{a_2}{a_1}. \tag{1.21}$$

This means that as a photon travels through an expanding Universe, it actually loses energy. Because a photon's energy is proportional to its frequency, we observe this as a reddening of light, or a redshift.

The redshift, defined as a change in frequency, ν, is related to a_1 and a_2 by:

$$z = \frac{\nu_1 - \nu_2}{\nu_2} = \frac{a_2}{a_1} - 1. \tag{1.22}$$

In a flat Universe, we can set $a = 1$ today[5], we have the useful relationship between the scale factor and the redshift of any object observed in the Universe:

$$a = \frac{1}{z+1}. \tag{1.23}$$

Often this redshift is interpreted as a Doppler shift in the the observed frequency of light. However, [9] would argue that this interpretation is incorrect. Consider the following (rather contrived) scenario. Imagine there are two galaxies in a Universe with a constant scale factor (the Universe is not expanding). Galaxy A emits a photon towards galaxy B. Just as the photon is emitted, the Universe begins to expand but when it is received in galaxy B, the Universe stops expanding so both galaxies are "at rest". Clearly, the photon will be redshifted due to the expansion of space during its journey, however at the time it was emitted and received,

[5] In general, $1 + z = a_0/a$ and a_0 is determined by the curvature. See [18].

the galaxies had zero velocity. Thus cosmological redshift has more to do with gravity than velocities.

Nevertheless, it is often useful to think of gravitational redshift as a Doppler shift, at least for nearby galaxies, and to consider the redshifted galaxies as receding from us, the observer, with velocity $v = cz$. We can determine a relationship between velocity and distance for sufficiently short distances.

From Eq.(1.13), the instantaneous distance between us and a galaxy at comoving radial coordinate χ is

$$d_p(t) = a(t)R_0\chi. \tag{1.24}$$

Then, since in this description χ is *comoving*, it remains constant as the Universe expands. Thus, the "velocity" is

$$v = \dot{d}_p = \dot{a}R_0\chi = \frac{\dot{a}}{a}d_p, \tag{1.25}$$

which, when evaluated today gives us the famous Hubble law [7]:

$$v = H_0 d_p. \tag{1.26}$$

H_0 is the current value of the time evolving Hubble parameter, defined as:

$$H = \frac{\dot{a}}{a}. \tag{1.27}$$

H_0 is also often parameterised as $100h\,\mathrm{km/s/Mpc}$, where the current value for h is around 0.67 [19]. H_0 also defines the Hubble length, d_H, which sets the characteristic scale for the Universe:

$$d_H = H_0^{-1}c. \tag{1.28}$$

This is the distance at which galaxies are receding from us at the speed of light and thus defines the minimum distance between regions of space which are causally disconnected. It should be noted that as H is time evolving, the Hubble radius also changes with time and so regions which were once causally connected may later become disconnected and vice versa (see Section 1.6.1).

The Hubble parameter is related to the density parameter, Ω, defined as [14]:

$$\Omega = \frac{8\pi G}{3H^2}\rho = \frac{\rho}{\rho_{\mathrm{crit}}}, \tag{1.29}$$

where

$$\rho_{\text{crit}} = \frac{3H^2}{8\pi G}. \tag{1.30}$$

This last equation is referred to as the critical density because it is the density at which the Universe is exactly flat. This can be seen by putting the definition for ρ_{crit} into the second Friedmann equation (Eq.(1.20)), which shows that $\kappa = 0$ if and only if $\rho = \rho_{\text{crit}}$.

We can rewrite Eq.(1.19) as:

$$H^2 = \frac{8\pi G}{3} \sum_i \rho_i, \tag{1.31}$$

which sums over all the energy densities of the components in the Universe. Note, here we write a fictitious energy density for curvature, which is not a true energy density but is useful in describing the behaviour of curvature in the Universe [9].

1.4 Distances in the Universe

The luminosity distance is defined as [20]:

$$d_L^2 = \frac{L}{4\pi F}, \tag{1.32}$$

where L is the luminosity of the source and F is the flux measured by the observer. Including the expansion of space, the cosmological equation for luminosity distance is [9, 20]:

$$d_L(z) = (1+z)\frac{c}{H_0\sqrt{-\Omega_k}} \sin\left(\sqrt{-\Omega_k}\int_0^z \frac{H_0}{H(z')}\mathrm{d}z'\right), \tag{1.33}$$

where the reader is reminded that this single equation is equivalent to the three separate curvature cases often quoted in textbooks, since $\sinh(x) = -i\sin(ix)$ [20]. Here, $\Omega_k = \kappa/(H_0^2 a^2)$ (see Section 1.5.4) is the curvature density parameter. For a flat Universe, Eq.(4.9) reduces to

$$d_L(z) = c(1+z)\int \frac{dz'}{H(z')}. \tag{1.34}$$

The angular diameter distance is defined as:

$$d_A = \frac{R}{\theta}, \tag{1.35}$$

where R is the proper size of an object and θ is its observed angular diameter. The distance duality relation, which holds in any metric theory of gravity where photon number is conserved [20], relates luminosity distance and angular diameter distance:

$$d_L(z) = (1+z)^2 d_A(z). \tag{1.36}$$

To relate distances to actual observations, we must consider magnitude. Magnitude is a (very old) astronomical measure of how much dimmer an object appears to the eye in comparison to some reference object, usually the star Vega. The apparent magnitude, m, of two stars is related to their measured fluxes, F, by [21]:

$$m_1 - m_2 = -2.5\log_{10}(F_1/F_2). \tag{1.37}$$

The absolute magnitude, M, of an object is the magnitude it would have if it were at a distance of 10pc. Thus, the distance modulus (the difference between the apparent and absolute magnitude of an object) is related to the luminosity distance by [21]:

$$\mu = m - M = 5\log_{10}\left(\frac{d_L}{10\text{pc}}\right). \tag{1.38}$$

Figure 1.2 shows a Hubble diagram using recent supernova data (see Section 1.7.1). Distance modulus measurements can be used to constrain cosmological parameters through Eq.(1.33).

1.4.1 Determining distances

Eq.(1.33) is, fairly obviously, a strong function of redshift. If one is willing to assume a cosmological model, then all that is needed to determine the distance to an object is the redshift.

The redshift of an object is usually determined from its spectrum (see Figure 1.3 for an example), although a less accurate, photometric redshift can be estimated from multi-band imaging [20]. From Eq.(1.22), it follows that:

$$z = \frac{\lambda - \lambda_0}{\lambda_0}, \tag{1.39}$$

where λ is the observed wavelength and λ_0 the wavelength at which the light was emitted.

Distances to the farthest objects in the Universe can only be estimated by utilising the distance ladder, which is illustrated in Figure 1.4 [3]. The idea is, the distance to only nearby objects can

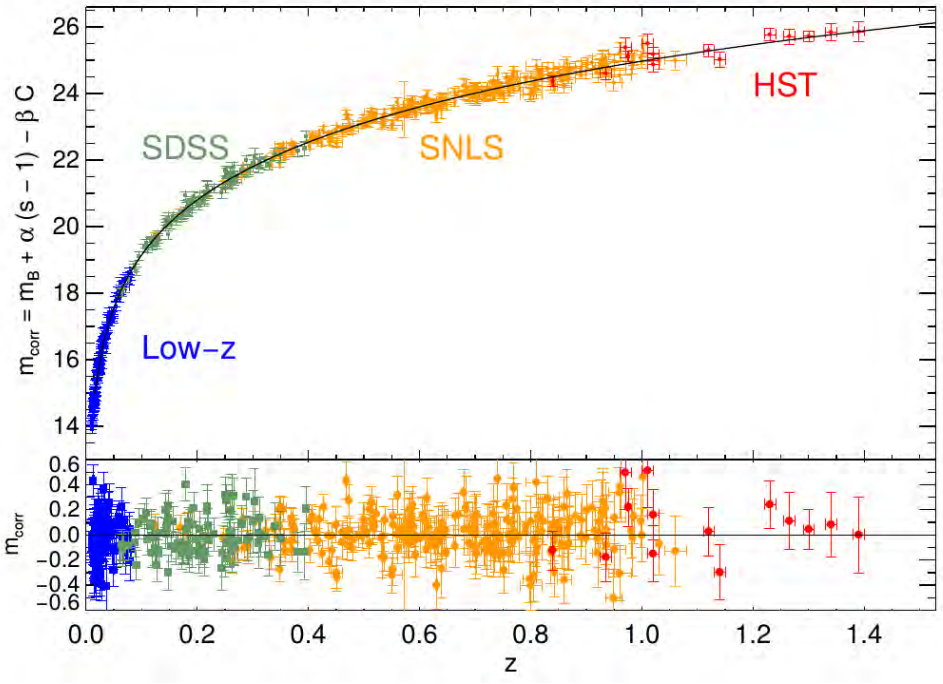


FIGURE 1.2: Hubble diagram of supernovae from various surveys showing the corrected magnitude as a function of redshift. In the lower panel, the best fit line is removed [22].

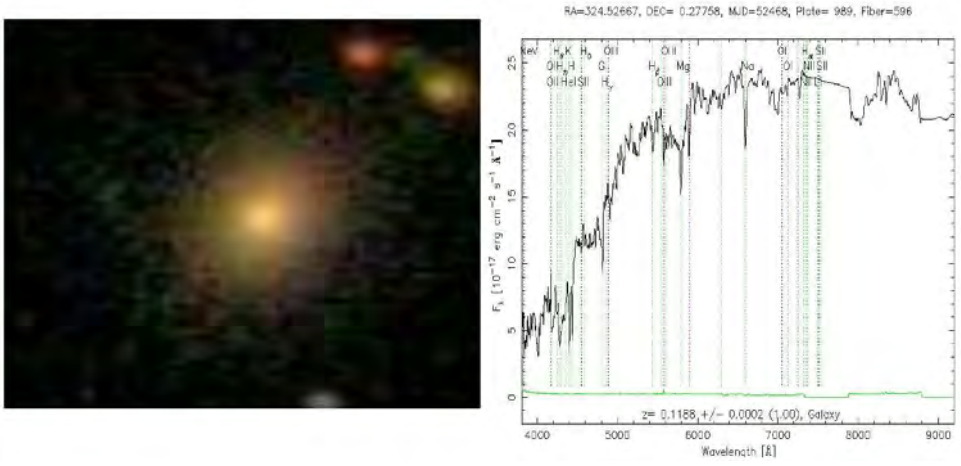


FIGURE 1.3: A typical spectrum of a luminous red galaxy, from which the redshift can be estimated using Eq.(1.39) [20].

be measured directly using parallax. For others, we need to use standard candles (and possibly standard rulers), objects of known brightness, to determine distances. The best example of these are Type Ia supernovae, which are discussed in Section 1.7.1.

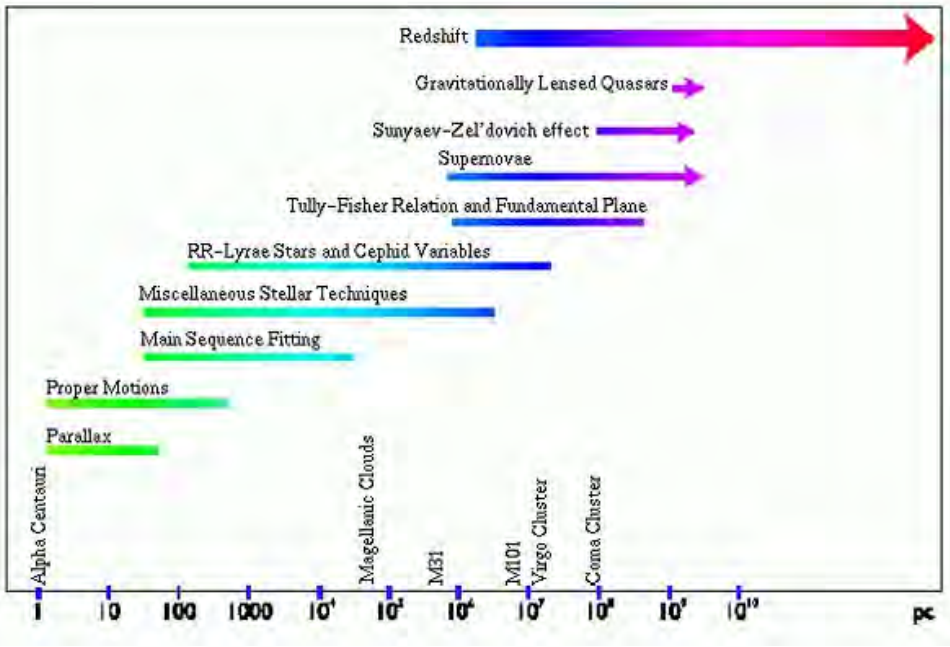


FIGURE 1.4: The cosmic distance ladder showing how different techniques are used to determine distances to objects. Parallax and proper motion are accurate techniques but only apply to nearby objects. Cepheid variables and RR lyrae variable stars are particularly useful for intermediate distances, while Type Ia supernovae have proven to be powerful probes at high redshifts. The region labeled "redshift" simply indicates that the distance can be determined to any object for which a redshift is measurable, provided a model of cosmology is assumed. Adapted from [23, 24].

1.5 Components of the Universe

1.5.1 Matter

It is now almost scientific consensus that there is a mysterious form of matter known as dark matter, which interacts with normal matter only gravitationally and has, thus far, only been detected through indirect means such as galaxy rotation curves and gravitational lensing [25]. Current constraints dictate that about 23% of the energy content in the Universe is dark matter and only 4% is normal baryonic matter, with a tiny fraction of radiation and about 73% dark energy (discussed below) making up the rest [8]. We assume that dark matter can be described as a pressureless dust, exactly the same as baryonic matter, which has thus far seemed consistent with observations. Thus we will group baryonic and dark matter together in our discussion of the evolution of the Universe [8].

Being pressureless, the equation of state for matter is zero: $w = 0$. Eq.(1.18) dictates that the energy density of matter evolves as:

$$\rho_{\mathrm{M}} \propto a^{-3}. \tag{1.40}$$

This can simply be interpreted as the decrease in the number density of matter particles as the Universe expands.

1.5.2 Radiation

Radiation can also be described as an ideal fluid. Radiation includes not only photons, but also massive particles moving at relativistic speeds which are moving fast enough to be indistinguishable from photons and thus have the same equation of state. Since $w = \frac{1}{3}$ for radiation, Eq.(1.18) tells us the energy density of radiation falls off as [3]:

$$\rho_{\mathrm{R}} \propto a^{-4}. \tag{1.41}$$

The extra factor of a^{-1} for radiation is because while the number density of photons decreases the same way as matter as the Universe expands, individual photons also lose energy as their wavelengths are redshifted, and the energy lost is proportional to the scale factor. Thus radiation dominates in the early Universe but is then overtaken by matter later on [3].

1.5.3 Dark Energy

In 1998, two independent teams of scientists discovered the first direct evidence that the expansion of the Universe was accelerating [26, 27] (see Section 1.7.1). Einstein's equation (Eq.(1.1)) can be modified in two ways to include the acceleration of the expansion of space. The first is to modify the left hand side of the equation, as done in Eq.(1.12). This then indicates a change in the nature of gravity on large scales. For example, the Ricci scalar, R, in Eq.(1.1) can be replaced by an arbitrary function $f(R)$, to capture characteristics of higher order gravity and provide an alternative to GR [28]. The other option is one can change the energy-momentum tensor on the right hand side of the equation to include an extra component, known as "dark energy", which causes the accelerated expansion [8].

In order for the expansion of the Universe to accelerate, we must have $\ddot{a} > 0$. Thus, using Eq.(1.20), we find $\rho + 3p < 0$. If we assume that dark energy follows an ideal gas law, like radiation and matter, then this implies that

$$w_{DE} < -\frac{1}{3}, \tag{1.42}$$

where $w_{DE} = \frac{p}{\rho}$ is the assumed equation of state for dark energy. This means that dark energy is a substance with negative pressure [8].

The simplest solution for the behaviour of dark energy is that its density does not change in time. This corresponds to adding a cosmological constant, Λ, in Eq.(1.12). Then the density parameter (today) for the cosmological constant is [29]:

$$\Omega_\Lambda = \frac{\Lambda}{3H_0^2}. \tag{1.43}$$

So the energy density of dark energy, if it is a cosmological constant, is constant in time:

$$\rho_\Lambda = \text{constant}. \tag{1.44}$$

Eq.(1.18) dictates that for the energy density to be constant, $w_\Lambda = -1$.

The cosmological constant model of dark energy, presents several difficult problems. One is known as the coincidence problem. That is, if the early Universe was matter dominated and the future will clearly be completely dominated by dark energy (since its energy density is constant and that of matter decreases as the Universe expands), then why do we happen to live at a seemingly special time when the energy density of matter and dark energy are the same order of magnitude? Proposed solutions to the coincidence problem include anthropic reasoning [30, 31] and tracking solutions using dynamical dark energy [32]. There is currently no consensus on the solution to the coincidence problem.

The second problem is why the measured value of the cosmological constant is so small. The source of the acceleration of the Universe has often been thought to be the energy of the quantum vacuum. All quantum fields fluctuate according to Heisenberg's uncertainty principle [33] and this results in a predicted non-zero energy density of the vacuum. The problem is that the theoretical prediction is stronger than the measured value of the quantum vacuum by $\sim 10^{120}$ [29]. The discrepancy between the measured and predicted quantum vacuum values is an outstanding problem in physics and has led scientists to seek alternative explanations for dark energy.

Although a cosmological constant model is the simplest which fits the data, we can also imagine dark energy whose properties vary in time and space. This idea of dynamical dark energy can be implemented simply using scalar fields and is also known as quintessence [32]. Observations

are have been unable to detect any evidence for dark energy dynamics and subsequently, modern dynamical models often mimic a cosmological constant model at the sensitivity of current observations [32].

1.5.4 Curvature

Curvature is an intrinsic property of space. Because we find it somewhat challenging to think of curvature of a three dimensional surface, it helps to visualise curvature of two dimensional surfaces when discussing curvature of the Universe. There are three possible curvatures of spacetime which are isotropic and homogeneous on large scales (in keeping with the cosmological principle): positive (sometimes referred to as closed), negative (sometimes referred to as open) and flat [2]. These are illustrated in Figure 1.5.

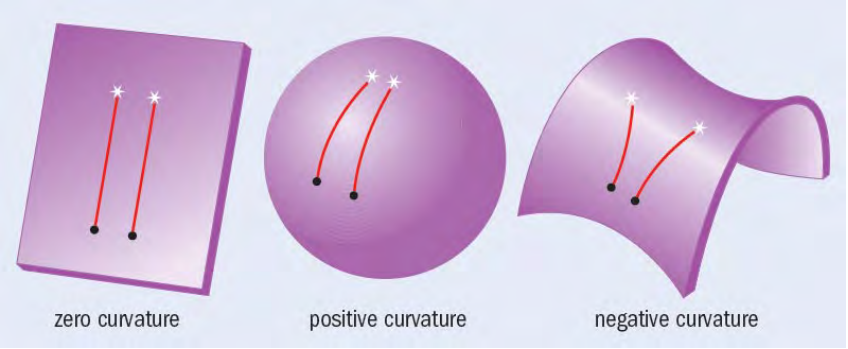


FIGURE 1.5: The three possible geometries of the Universe: spherical ($\Omega_0 > 1$), hyperbolic ($\Omega_0 < 1$) and flat ($\Omega_0 = 1$) where Ω_0 is the total density parameter [34].

In order to include curvature neatly in Eq.(1.31), we can construct an energy density for curvature which, from Eq.(1.19), must be [9]:

$$\rho_k = \frac{3\kappa}{8\pi G a^2} \tag{1.45}$$

and a corresponding density parameter:

$$\Omega_k = \frac{\kappa}{H_0^2 a^2}. \tag{1.46}$$

Flat space is familiar Euclidean space, where the angles of triangles add up to 180°, the circumference of a circle is $2\pi r$ and parallel lines remain parallel forever. In two dimensions, a flat Universe could be represented by an infinite sheet of paper. A flat Universe must be infinite because otherwise the boundaries would constitute a special place and violate isotropy. In a flat

Universe, the density exactly equals the critical density (Eq.(1.30)), which makes the curvature density parameter equal to zero [14]:

$$\Omega_k = 0. \tag{1.47}$$

Positive curvature is represented by the surface of a sphere in two dimensions. In spherical geometry, as on the surface of the Earth, the angles of a triangle will add up to a number greater than 180°, the circumference of a circle is less than $2\pi r$ and lines which start parallel will end up meeting (lines of longitude for example). Spherical geometry describes a three-sphere, meaning it has a finite size but no boundary while still maintaining isotropy (like the surface of the Earth). In a spherical Universe, the curvature density parameter is positive [14]:

$$\Omega_k > 0. \tag{1.48}$$

Lastly, hyperbolic geometry (negative curvature) can be visualised by a saddle. Here all angles of a triangle add up to less than 180°, the circumference of a circle is greater than $2\pi r$ and lines which are initially parallel diverge from each other and never meet. In a hyperbolic Universe, the curvature density parameter is negative [14]:

$$\Omega_k < 0. \tag{1.49}$$

This means that for curvature, $w = -\frac{1}{3}$ and $\rho_k \propto a^{-2}$ [9].

Finally, it should be noted that Eq.(1.7) completely describes mathematically the curvature of a manifold, which implies that all information about curvature is contained in the metric [9].

1.5.5 The Friedmann Equation

Now we can rewrite Eq.(1.31) in terms of density parameters. We first recognise that

$$\frac{8\pi G}{3} = \frac{H_0^2}{\rho_{\text{crit},0}}, \tag{1.50}$$

where $\rho_{\text{crit},0}$ is the value of the critical energy density today. We can then input the evolution of each component of the first Friedmann equation, and write it in terms of density parameters, as they are measured today [2]:

$$H^2 = H_0^2\big(\Omega_m(1+z)^3 + \Omega_r(1+z)^4 + \Omega_k(1+z)^2 + \Omega_\Lambda\big). \tag{1.51}$$

1.6 Problems with the Big Bang

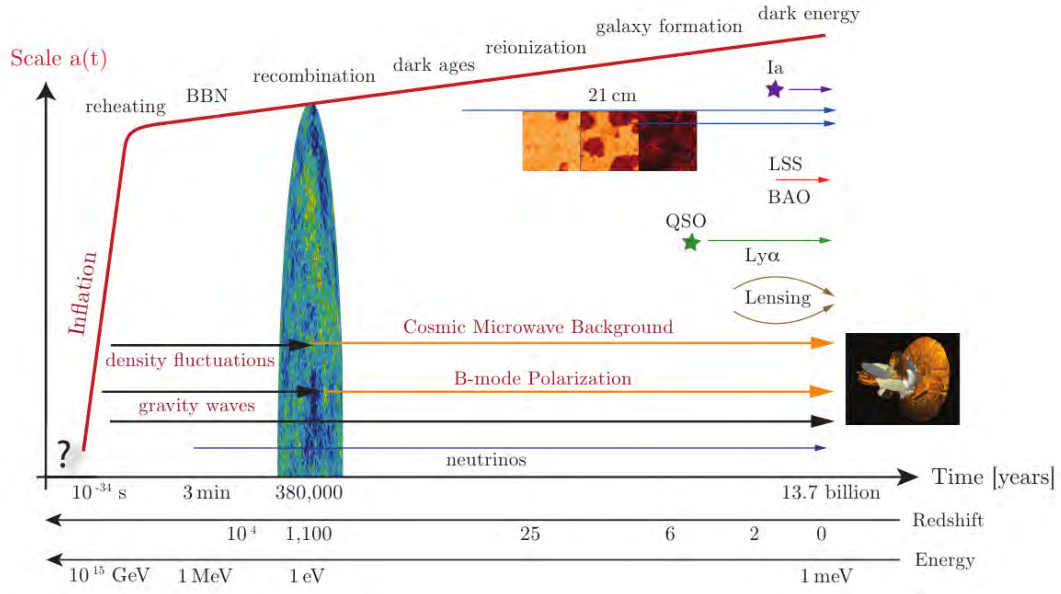


FIGURE 1.6: The standard model of the history of the Universe showing important events as a function of time, redshift and energy scale. Also indicated are the observations which probe this time scale. The acronyms are: BBN (Big Bang Nucleosynthesis), Ia (Type Ia Supernovae), LSS (Large Scale Structure), BAO (Baryon Acoustic Oscillations), QSO (Quasi-Stellar Objects), Lyα (Lyman-Alpha) and 21 cm (hydrogen 21 cm transition) [35].

Figure 1.6 represents our current understanding of the history of the Universe, along with the observations which probe the various epochs. Modern particle accelerators are powerful enough to probe energy scales up to about 1TeV, which corresponds to when the Universe was just 10^{-10} seconds old. At these energies, we can be confident we understand the physics which dictates the evolution of the Universe. What happened prior to this, however, is the realm of theory and hypothesis. The development of a theory of the early Universe has been largely driven by the need to explain the initial conditions of the Big Bang model [2].

There are two main problems in the standard Big Bang Theory which requires extremely fine-tuned initial conditions. These are the horizon and flatness problems.

The horizon problem

From the CMB [3] we know that the early Universe was highly homogeneous. Yet we would expect regions of the early Universe to be causally disconnected. There is no reason why these causally disconnected regions should have similar physical conditions. Thus the standard Big Bang Model requires the initial distribution of matter to be fine-tuned to be extremely smooth [35].

The flatness problem

To illustrate the flatness problem, we begin by rewriting Eq.(1.19) as

$$1 - \Omega(a) = -\frac{\kappa}{(aH)^2}. \tag{1.52}$$

where $\Omega(a)$ is the total energy density in the Universe. Since the comoving Hubble radius, (aH^{-1}), grows with time, we see that $|1 - \Omega|$ must diverge. Thus in order for Ω to be close to one today (as data suggests), Ω must have been extremely close to one in the early Universe, making the initial conditions once again highly fine-tuned [8].

1.6.1 Inflation

These two shortcomings of the standard Big Bang model are problems with initial conditions, in that the model is only consistent with current observations if the Universe is fine-tuned to initially be extremely flat and smooth. Additionally, the standard Big Bang model cannot explain how the seeds for large scale structure formation are generated, these initial conditions must also be assumed.

The theory of inflation [36] solves the flatness and horizon problems and explains where the seeds for structure formation came from. The idea of inflation is simple: in the early Universe there was a period of rapid acceleration, corresponding to a decreasing comoving Hubble radius. From Eq.(1.52), we can see that if the comoving Hubble radius, $(\mathrm{aH})^{-1}$, decreases with time then Ω is driven towards 1: a flat Universe. The horizon problem is also solved because what are large scales today were in fact within the Hubble radius before inflation, allowing causal connections to explain homogeneity [3].

The simplest models of inflation state that a single scalar field, the inflaton, was responsible for the rapid acceleration during inflation. Figure 1.7 illustrates a simple potential for such a scalar field. The acceleration of the scale factor can be related to the first time derivative of the Hubble parameter [35]:

$$\frac{\ddot{a}}{a} = H^2(1 - \epsilon), \tag{1.53}$$

where $\epsilon = -\frac{\dot{H}}{H^2}$. Thus, the condition for acceleration is:

$$\epsilon = -\frac{\dot{H}}{H^2} = -\frac{d\mathrm{lnH}}{dN} < 1, \tag{1.54}$$

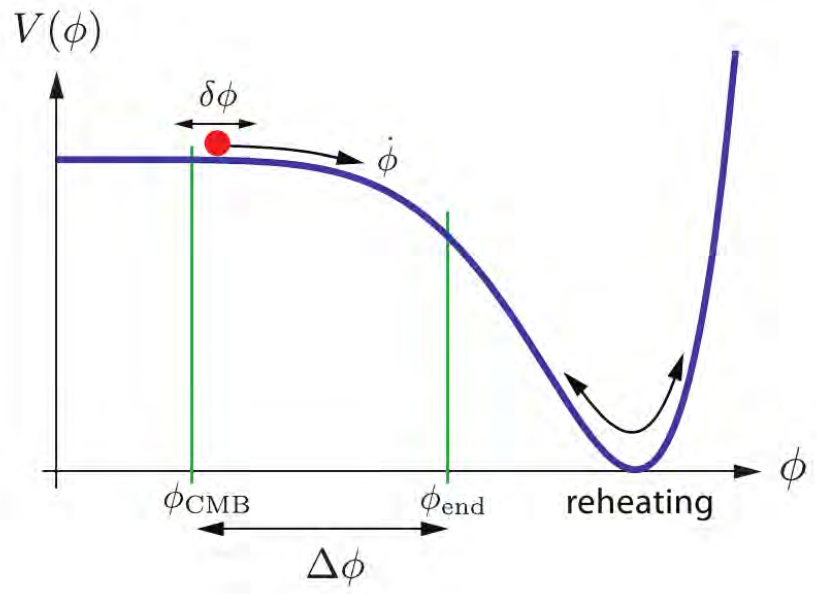


FIGURE 1.7: Example of a simple inflaton potential. Acceleration occurs while the potential energy, V(ϕ), dominates over the kinetic energy, $\frac{1}{2}\dot{\phi}^2$. The quantum fluctuations, $\delta\phi$, which form the seeds for structure growth and are visible as the CMB, occur at $\phi_{\rm CMB}$, about 60 e-folds (see definition below) before the end of inflation. Inflation ends at $\phi_{\rm end}$ when the kinetic energy is comparable to the potential energy [35].

where we have defined $dN = Hdt = d\ln a$, which measures the number of e-folds, N, of expansion. One e-fold would mean the Universe had expanded by a factor of e. This equation therefore shows that in order to have acceleration, the fractional change of the Hubble parameter per e-fold must be small.

Finally, from the action of the potential in Figure 1.7 it is possible to derive the equation of state for the inflaton field, assuming it is a perfect fluid [35]:

$$w_\phi = \frac{p_\phi}{\rho_\phi} = \frac{\frac{1}{2}\dot{\phi}^2 - V}{\frac{1}{2}\dot{\phi}^2 + V}, \tag{1.55}$$

which illustrates how a scalar field can lead to negative pressure and accelerated expansion ($w_\phi < -\frac{1}{3}$) if the potential energy dominates over the kinetic.

Inflation is a successful theory of the early Universe not only because it can explain why the Universe is so flat and smooth, but also because it provides a natural source for the seeds of structure formation: quantum fluctuations in the scalar field. Due to Heisenberg's uncertainty principle (specifically, the uncertainty relation between time and energy), energy and particles are continually popping in and out of existence on short time scales [2, 3]. During inflation, the quantum fluctuations in the inflaton field, $\delta\phi$, get stretched out as the Universe rapidly expands.

These fluctuations in the scalar field lead to local time delays in the end of inflation, meaning that some parts of the Universe will undergo inflation slightly longer than other parts. This in turn results in density fluctuations throughout the entire Universe. It is these density fluctuations, these initial seeds, which gravitational attraction eventually turns into galaxies and the large scale structure we observe today. The fact that the measured power spectrum of these fluctuations from the CMB [3] matches the one predicted by inflation so precisely provides the strongest evidence we have for inflation [35]. Of course, this simple model for inflation is just one of a dazzling variety of models, a study of which is out of the scope of this thesis.

1.7 Observations of the Universe

Having set the cosmological stage, this section serves as an overview of specific fields of observational cosmology and astronomy relevant to later chapters.

1.7.1 Supernova Cosmology

Supernovae are among the most powerful events in the Universe. They occur when a star dies, causing an explosion so massive it can temporarily outshine an entire galaxy. Supernovae (SNe) are classified into several types, based on their spectral features: Type I supernova have no hydrogen in their spectra, while Type II do. Type Ib have helium, Type Ic have no hydrogen or helium and Type Ia also lack hydrogen and helium but are dominated by lines from higher-mass elements (such as calcium, silicon, sulphur and iron). A simple table is also shown in Figure 1.8 to summarise the difference between the types and a spectrum of a supernova of each of these types are illustrated in Figure 1.9. Over the years, a host of objects have been discovered which do not fit into the standard classification, which is constantly being revised [37].

It is believed that Type II and Type Ib/c SNe are due to the core-collapse of massive stars. They seem to leave behind either a remnant neutron star or black hole, depending on the initial mass of the collapsing star. There is much variation in the light curves of these supernovae (see Figure 1.10). Type Ia supernovae, however, are far more consistent. Type Ia SNe are probably caused by the thermonuclear explosion of a white dwarf star. The white dwarf accretes mass from a companion (either a normal star or by merging with another white dwarf) until it reaches the Chandrasekhar mass ($\sim 1.4 M_{\odot}$), triggering a runaway reaction burning all the carbon in the star within seconds. The light we observe lingering after a supernova explosion comes from

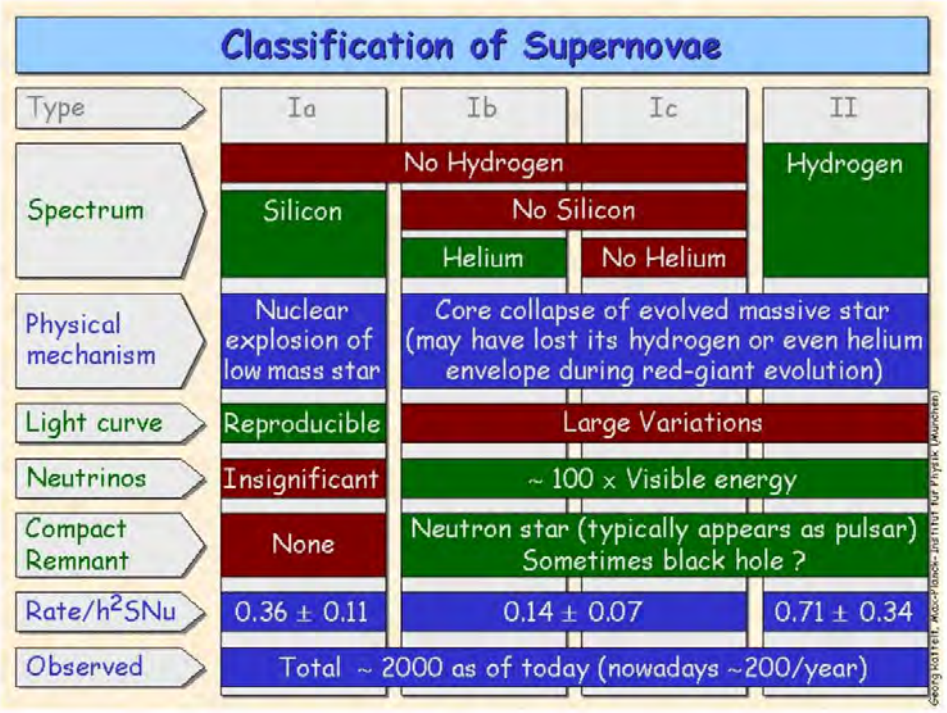


FIGURE 1.8: The spectral classification of supernovae. The type of a supernova is decided based on a decision tree on the existence of certain spectral lines. This table shows that Type Ia's are unique in that their light curves show consistency, amongst other features. The rate is measured in supernova units, where 1 SNu = 1 SN per century per $10^{10} L_{\odot,B}$ (the B-band solar luminosity) [38].

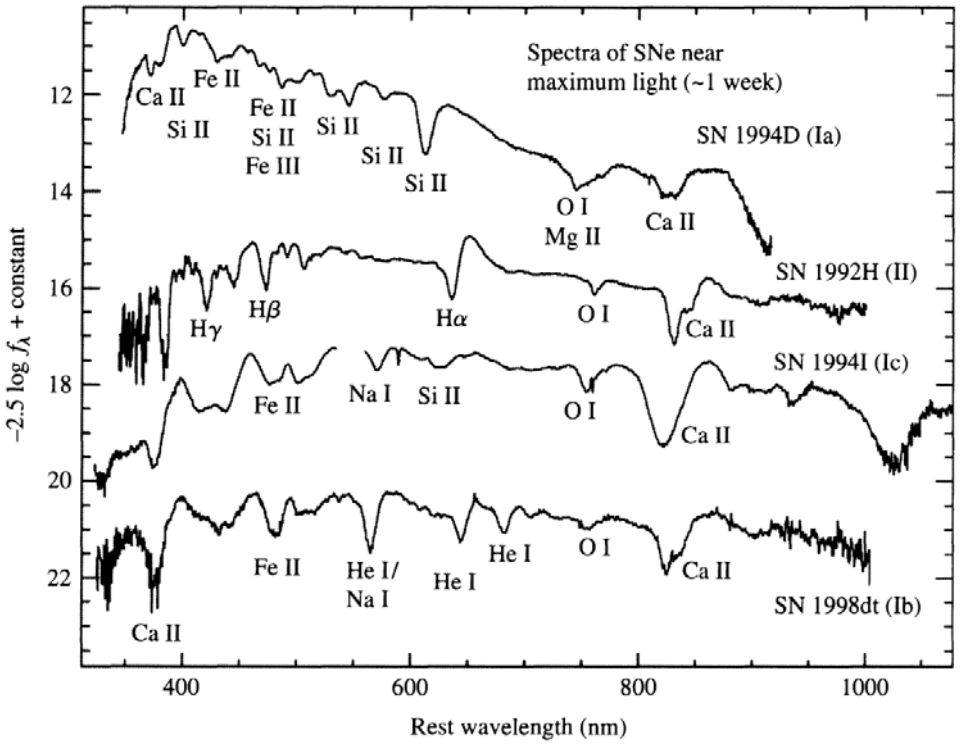


FIGURE 1.9: Representative spectra of different supernova types. The type is in brackets after the supernova name. Spectral lines are indicated by their element name and ionization. The brightness is in arbitrary flux units [39].

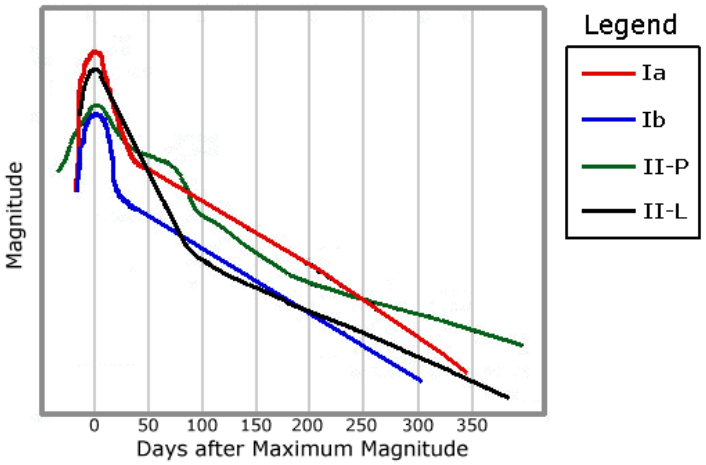


FIGURE 1.10: Examples of the differences between the types of supernovae light curves. While the differences are not distinct enough to accurately type supernovae as with their spectra, light curves can still give some indication as to the probability the supernova is a Type Ia for example (see Chapter 4) [40].

radioactive decay of new material synthesised in the explosion, mostly nickel decaying into iron [37].

Because Type Ia SNe are so similar in their light curves and spectra, they can be used as "standard candles": objects of known brightness which can be used to determine cosmological distances. In fact, Type Ia SNe are "standardisable candles" because each light curve is not identical, but there is a tight relationship between the absolute brightness of the object and the stretch and colour of the light curve. Figure 1.11 shows an example of Type Ia light curves before and after stretch and colour correction. Other correction terms to the magnitude of the supernova include dust extinction and a K correction term (needed to transfer the observed flux into the rest frame) [37].

Type Ia SNe are excellent cosmological probes, especially of H_0 and Ω_m. In the nineties, Perlmutter *et al.* [27] and Riess *et al.* [26] used their independent measurements of supernovae to find the first evidence for dark energy. Figure 1.12 shows the confidence limits on the cosmological parameters, Ω_m and Ω_Λ, showing how a cosmological constant of zero is ruled out at high confidence by the supernova data. Also illustrated is how the cosmological probes, supernovae, the CMB and the BAO, are extremely complementary[6].

[6] Also note the remarkable amount the contours have shrunk in little over a decade.

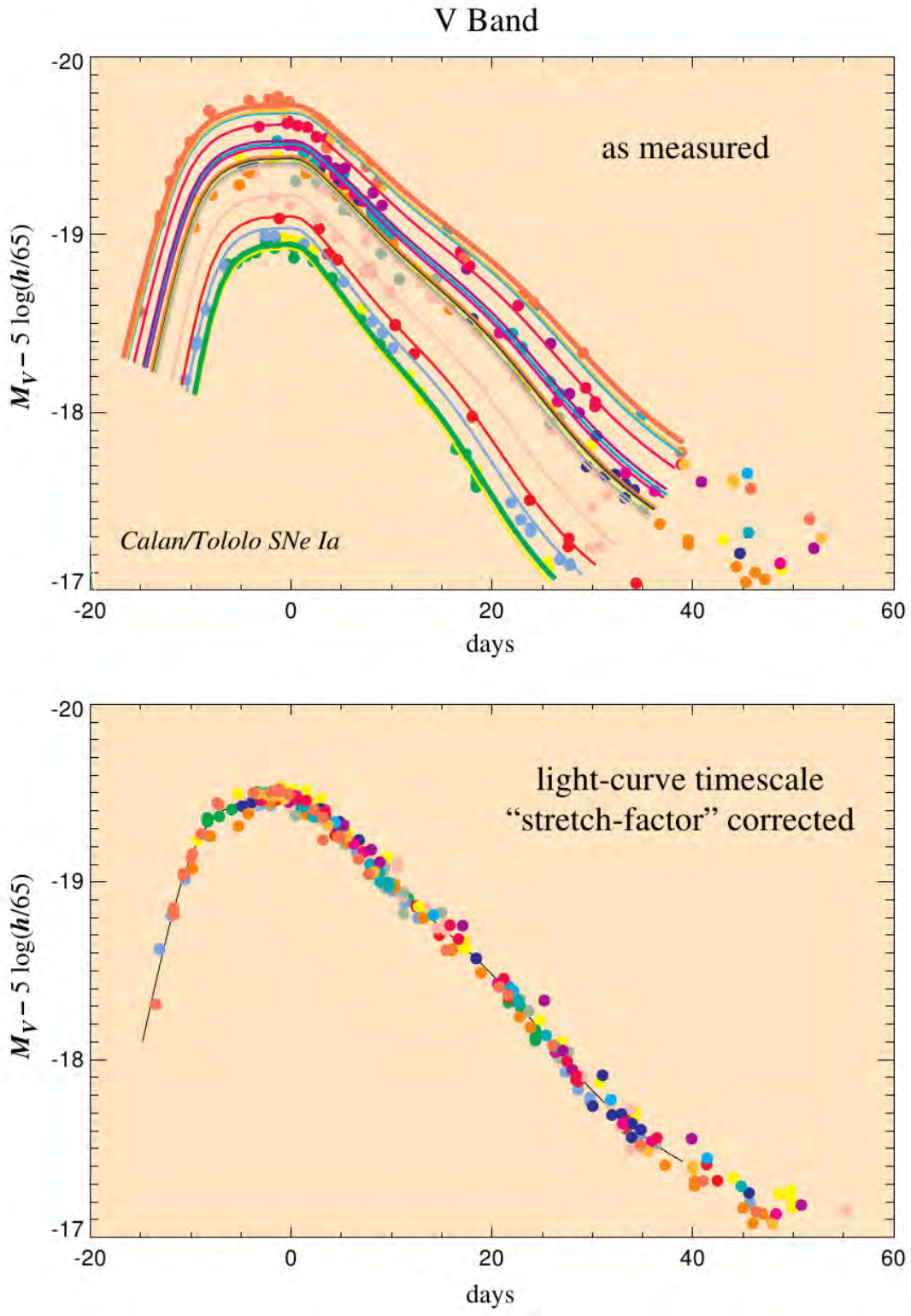


FIGURE 1.11: Example light curves (brightness as a function of time) of nearby supernovae. These show the relationship between peak absolute luminosity and the timescale of the light curve: brighter supernovae decay more slowly than fainter ones. The bottom panel shows the same light curves as the top panel after fitting and removing the stretch factor thereby "correcting" the peak magnitude and aligning the light curves [41].

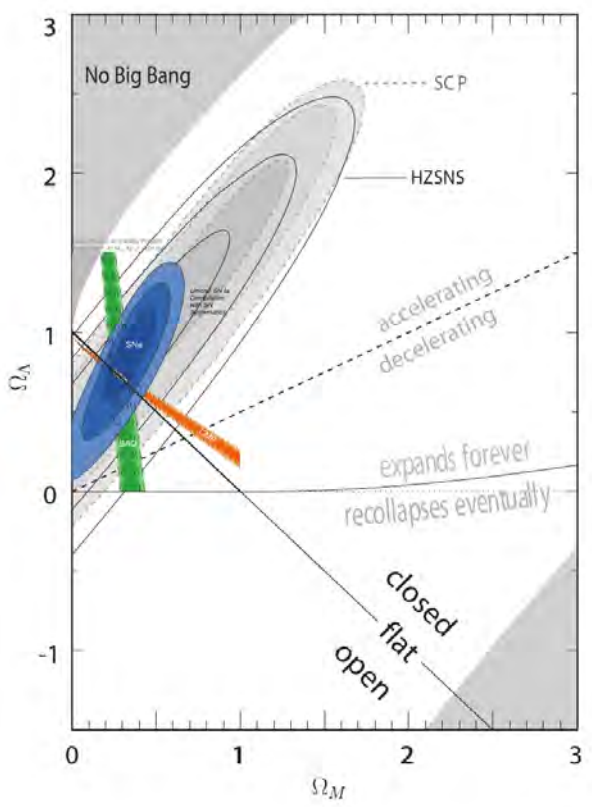


FIGURE 1.12: The $\Omega_m - \Omega_\Lambda$ plane showing the original supernova constraints from the High-Z Supernova Survey [26] and the Supernova Cosmology Project [27], as well as from the recent Union2 SNe Ia compilation [42]. Constraints from the CMB [43] and BAO [44] are also included [37].

The field of supernova cosmology is well established and is continuing to develop. Surveys such as DES[7], Pan-STARRS[8] and the LSST[9] will produce an unprecedented amount of supernova data in the coming years. However, such a deluge of data is riddled with difficulties when standard analysis techniques are applied. Thousands of supernovae will be discovered by these surveys and only a fraction will be confirmed spectroscopically as Type Ia. Either the majority of the data must be discarded, or a certain amount of contamination from non-Ia supernovae will enter the dataset, potentially leading to biases in the estimates of the cosmological parameters. In Chapter 4, we review these problems and an elegant statistical technique called BEAMS [45–47], Bayesian Estimation Applied to Multiple Species, which solves them. BEAMS can produce tight unbiased constraints on the cosmological parameters without applying any cuts to the dataset. We also develop in Chapter 4, a method of extending BEAMS to correlated datasets [1], since it is known that correlations between data will be important in these large surveys.

[7]The Dark Energy Survey, http://www.darkenergysurvey.org/
[8]The Panoramic Survey Telescope & Rapid Response System, http://pan-starrs.ifa.hawaii.edu/public/
[9]The Large Synoptic Survey Telescope, http://www.lsst.org/lsst/

1.7.2 The Epoch of Reionization

The period after recombination [3] is known as the dark ages, when the Universe was transparent and the first stars had not yet formed [48]. During this time, dark matter dominated the dynamics of the Universe, driving gravitational collapse and the beginnings of structure formation [8]. The dark ages ended after about 400 million years when the first galaxies had formed and the first stars (known as Population III stars) began to shine and emit ionizing ultraviolet radiation. The initially neutral intergalactic medium began to ionize around the source galaxies forming pockets of ionized hydrogen (HII). Gradually, as more galaxies formed, these pockets of ionized gas grew until they permeated the whole Universe and no neutral hydrogen remained [48].

Current observations place this Epoch of Reionization (EoR) in the redshift range $z \sim 6-15$ [48]. Figure 1.13 shows a simulated view of neutral hydrogen during the EoR. Before the EoR, the neutral hydrogen traces the cosmological density fluctuations of the early Universe [3]. However, as the source galaxies start to form, pockets of ionized hydrogen also form until at a redshift of about 6, no neutral hydrogen remains [48].

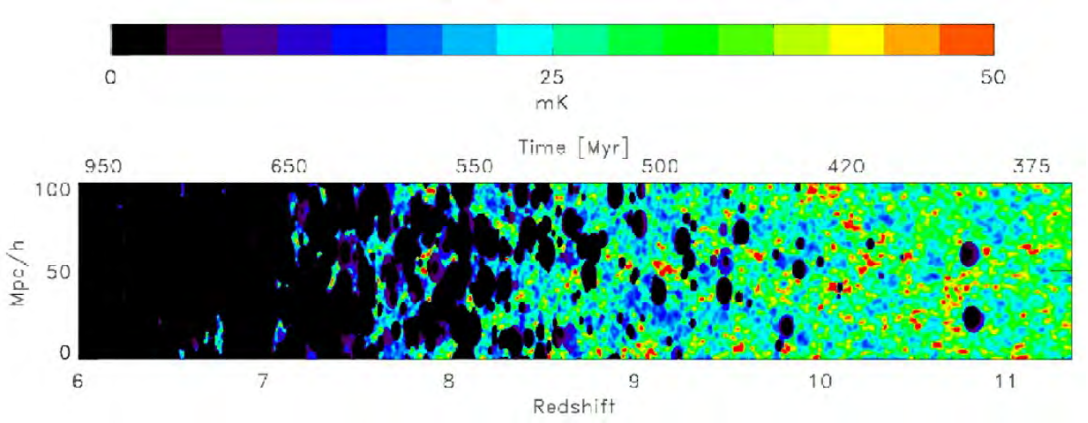


FIGURE 1.13: A simulation of neutral hydrogen during the Epoch of Reionization showing how pockets of ionized hydrogen form and eventually permeate the Universe by $z \approx 6$. The colour corresponds to the intensity of emission from neutral hydrogen [48].

The Epoch of Reionization is a difficult era to study and constraints on the details of the EoR are often weak and model dependent [48]. However, there are several types of observations which yield information on the EoR.

- **Lyman α forest** Ultraviolet light passing through neutral hydrogen is absorbed at a wavelength of 1215.67Å (in the rest frame), in the Lyman α transition. As the light from distant quasars travels towards Earth, it passes through clouds of neutral hydrogen. Since each cloud is at a different redshift, this forms the so-called Lyman α forest of spectral lines in the spectrum of the quasar, which is a probe of the intergalactic medium. At a redshift of about 6 there is so much neutral hydrogen that the forest becomes an absorption trough, called the Gunn-Peterson trough [49]. The fact that no trough is detected in quasars just below redshift of 6 but is in high redshift quasars is strong evidence that reionization did occur and ended at $z \sim 6$ [48].

- **Thomson scattering of the CMB** Reionization has an effect on the CMB photons through Thomson scattering [3]. The electrons from the ionized hydrogen scatter the CMB photons in random directions, which has the effect of washing out the small scale anisotropies and polarising the CMB photons. Studies of the polarisation of the CMB can constrain the optical depth of reionization, τ, and in turn (assuming a model of reionization history) the redshift at which reionization occurred [48].

- **The intergalactic medium** Measurements of the temperature of the intergalactic medium (as determined from the density) constrain the epoch of reionization to have occurred at $z \lesssim 10$ [48]. Studies of both the Lyman α forest and the number of galaxies detected optically at $z \gtrsim 7$ suggest that the number of ionizing photons per baryon may have been surprisingly small during the EoR.

- **The 21 cm line** Perhaps the most promising method to investigate the EoR is the neutral hydrogen emission line, with a rest frame wavelength of 21 cm. This emission occurs when the electron and proton of a neutral hydrogen atom transition from the state where the spins of each are parallel, to the lower energy anti-parallel state. Although this transition is extremely rare, the large amounts of hydrogen in the Universe makes this emission line possible, if difficult, to detect, even at high redshift [50].

  Current studies of the EoR are limited to detections of neutral hydrogen emission at $6 \lesssim z \lesssim 12$ because of ionisopheric interference at lower frequencies. Current and near future experiments such as GMRT[10], LOFAR[11], MWA[12] and PAPER[13] do not have the sensitivity to directly detect the reionization evolution. They instead plan on detecting

[10] Giant Metre Wave Telescope, http://gmrt.ncra.tifr.res.in
[11] Low Frequency Array, http://www.lofar.org
[12] Murchinson Widefield Array, http://www.mwatelescope.org
[13] Precision Array to Probe the Epoch of Reionization, http://astro.berkeley.edu/~dbacker/eor

the signal from the EoR statistically, by measuring the 21 cm power spectrum of variations in intensity and using this to constrain models of reionization.

The SKA, on the other hand, will have far greater sensitivity and resolution. A study of the EoR with the SKA low frequency array should be able to reveal the nature of the ionizing sources and could provide constraints on the non-gaussianity of the primordial cosmological density field [50]. The SKA, due to its large frequency range, will be able to probe up to $z \sim 30$ into the dark ages and will be able to actually map the EoR as it changes with redshift, instead of just detecting a signal statistically [48, 50]. With the combination of modeling foreground sources (a major source of interference), dealing with instrumental errors and the extraction of scientific parameters (the EoR signal) from data, this is exactly the kind of problem applicable to BIRO, the technique we develop in Chapter 5.

Other, less constraining probes of the EoR include cosmic infrared and x-ray background, Lyman α emitters, high redshift quasars, metal abundances at high redshift etc. [48]. These probes, along with the Lyman α forest, CMB polarisation and 21 cm studies, represent an exciting new era in understanding the Epoch of Reionization and the astrophysics of the high redshift intergalactic medium.

1.7.3 Large Scale Structure

Figure 1.14 shows the 2dF galaxy survey [51], demonstrating that the galaxies in our Universe are not distributed randomly but instead form colossal walls and voids, often referred to as the cosmic web. This large scale structure arises from the tiny seeds of overdense regions generated in the early Universe by inflation stretching out quantum fluctuations in the inflaton field (see Section 1.6.1). Gravitational instability, the process by which the attraction of gravity causes dense regions to become more dense, grows these seeds from tiny inhomogeneities into the pattern of massive clusters of galaxies and empty voids seen today [21].

The features of Figure 1.14 are slightly more exaggerated than they appear in physical space. The walls appear sharper and narrower than they actually are because the redshift of any galaxy moving towards the wall will be increased/decreased (depending on whether the galaxy is moving away or towards us) which sharpens the appearance of the wall if redshift is converted directly to distance. Similarly, clusters of galaxies appear elongated in the direction of the observer due

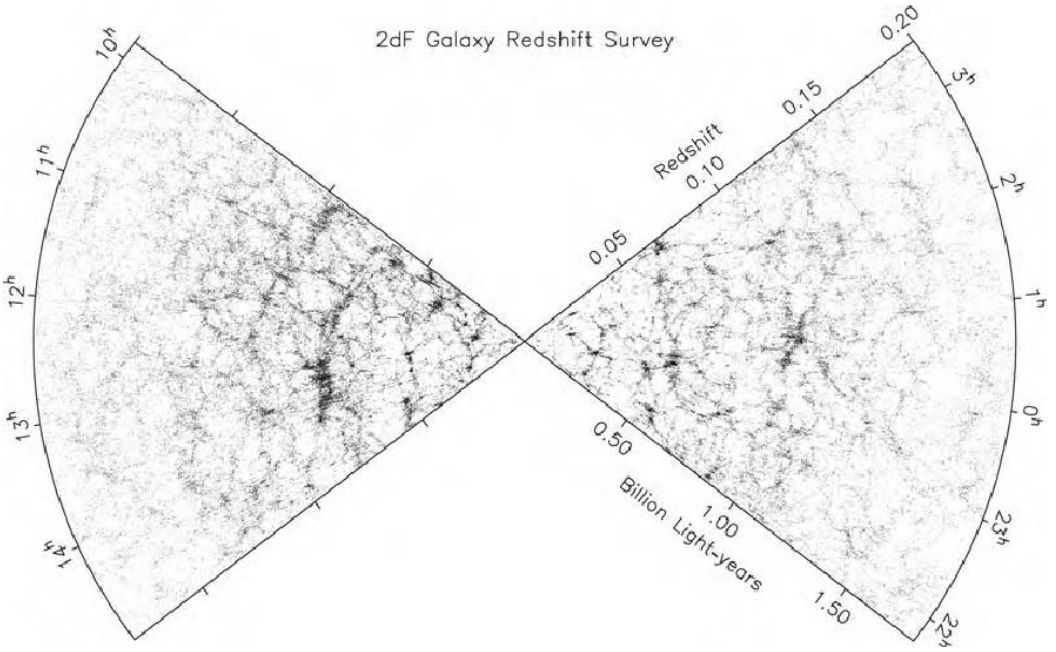


FIGURE 1.14: A slice of the galaxy distribution of the Universe from the 2dF survey. There are 93 170 galaxies in slices $-4° < \delta < 2°$ in the north (left wedge) and $-32° < \delta < -28°$ in the south (right wedge) [51].

to peculiar velocities distorting the redshift [21]. These elongated features are often referred to as "fingers of God".

The observed cosmic web evolved from smaller overdensities in the early Universe [8]. Objects such as stars exist in a state of continual battle between gravity trying to collapse them and radiation pressure holding them up. Only gas clouds with sufficient mass, the Jeans' mass [52], are able to beat the radiation pressure and collapse to form gravitationally bound objects [21]. The standard view is a bottom up approach: smaller objects such as galaxies formed first and then attracted each other to form galaxy clusters and the large walls and filaments seen in Figure 1.14. However, without dark matter we would not see this structure. Dark matter, since it is unaffected by radiation pressure, is able to form large objects known as halos, on the scale of galaxy clusters, which draw baryonic matter towards them and aid in structure formation.

One of the key science projects of the SKA is the study of galaxy evolution and cosmology with a large HI galaxy survey. The sensitivity of the SKA, coupled with the ubiquitousness of HI in the Universe, means this survey will be a powerful probe of how stars form from neutral gas, the role gas plays in galaxy dynamics as well as other questions in galaxy evolution. Figure 1.15 shows a simulation of the observing cone of the SKA and the objects it will be able to observe.

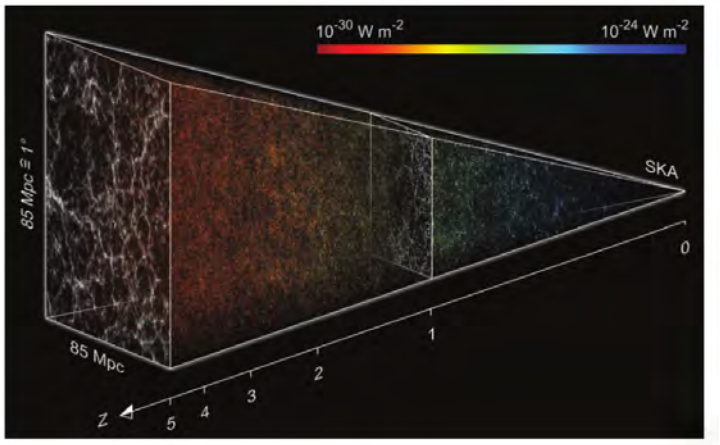


FIGURE 1.15: Simulated SKA HI observing cone. The SKA HI galaxy survey will be of far greater number and depth than any current galaxy survey [53].

1.7.3.1 Galaxy Clustering

The two-point correlation function, $\xi(r)$, is a useful tool for describing galaxy clustering. If the average spatial density of galaxies is n galaxies per cubic megaparsec, then the probability of finding a galaxy in a small volume, ΔV_1, is simply $n\Delta V_1$. If the galaxies are clustered we would expect, the probability of also finding a galaxy in a second small volume, ΔV_2, to be greater if the distance between the volumes, r_{12}, is small. The joint probability of finding a galaxy in both volumes is [21]:

$$\Delta P = n^2[1 + \xi(r_{12})]\Delta V_1 \Delta V_2. \tag{1.56}$$

Here, if $\xi(r) > 0$, the galaxies are clustered. If $\xi(r) < 0$ they tend to avoid each other and if $\xi(r) = 0$ there is no correlation between galaxies. A simple estimator for $\xi(r)$ is given by [54]:

$$1 + \xi(r) = \frac{DD(r)}{RR(r)}, \tag{1.57}$$

where $DD(r)$ is the number of pairs of galaxies in a catalogue, separated by distance r, and $RR(r)$ is the number of pairs of randomly distributed galaxies, separated by r. If $\xi(r) = 0$ at small r, then the clustering is no more than random.

A more sophisticated estimator, which has a lower variance, also utilises the galaxy pairs between the random and data catalogue, $DR(r)$, and is given by [54]:

$$1 + \xi(r) = \frac{DD(r) - 2DR(r) + RR(r)}{RR(r)}. \tag{1.58}$$

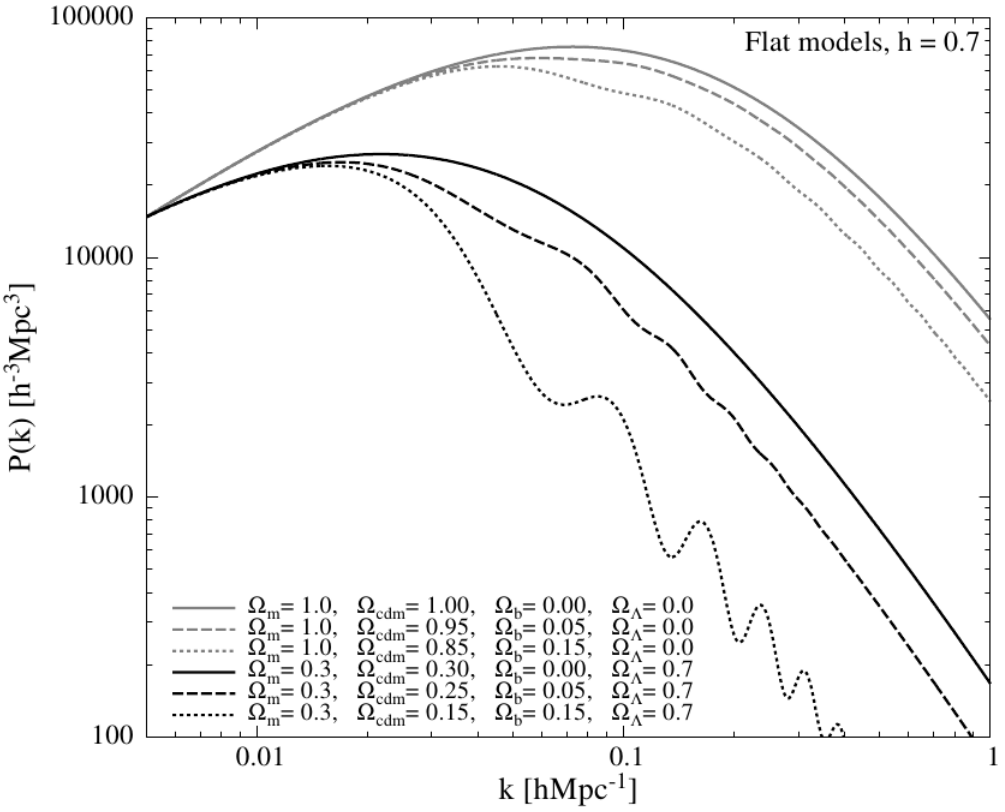


FIGURE 1.16: The effect of changing cosmological parameters on the matter power spectrum, assuming a flat Universe. Ω_m, Ω_{cdm}, Ω_b and Ω_Λ are the density of matter, cold dark matter, baryons and dark energy respectively [55].

The Fourier transform of the two-point correlation function is the power spectrum, $P(k)$ [21]:

$$P(k) = \int \xi(\mathbf{r}) \exp(i\mathbf{k} \cdot \mathbf{r}) \mathrm{d}^3\mathbf{r} = 4\pi \int_0^\infty \xi(r) \frac{\sin(kr)}{kr} r^2 \mathrm{d}r, \tag{1.59}$$

where bold symbols indicate vectors and k is the wavenumber, such that small k represents large scales. Figure 1.16 illustrates the effect the cosmological parameters have on the matter power spectrum and hence how it can be used to constrain the parameters. The oscillations in the power spectrum are produced by Baryon Acoustic Oscillations (BAO) in the early Universe (see Section 1.7.3.2).

Another way to consider the clustering of galaxies is in the the deviation from the average density. We can write the density at position $\mathbf{x}$ as $\rho(\mathbf{x}) = \bar{\rho}\ [1 + \delta(\mathbf{x})]$, where $\bar{\rho}$ is the mean density. Then, we define δ_R as the fractional deviation $\delta(\mathbf{x})$ averaged within a sphere of radius R. The average over all spheres of radius R of δ_R will be zero but the variance, σ_R^2, is a measure of how clumpy the galaxy distribution is on this scale. The variance is related to the power

spectrum by [56]

$$\sigma_R^2 = \frac{1}{8\pi^3}\int W^2(kR)P(k)d^3k, \tag{1.60}$$

where $W(kR)$ is the filter function (a top-hat function for a sphere of radius R) in Fourier space. Clustering is often parameterised by σ_8, the variance on scales of $8h^{-1}$Mpc, which is estimated to be around 0.9 [21].

1.7.3.2 Baryon Acoustic Oscillations

One of the most useful cosmological probes is the Baryon Acoustic Oscillations signature in the matter power spectrum. The BAO form a "standard ruler" and thus an excellent distance measure for the Universe. Before decoupling [3], the constant interplay between radiation pressure and gravity set up oscillations in the hot plasma of the early Universe. Any given perturbation will propagate at the speed of sound [20],

$$c_s = \frac{c}{\sqrt{3(1+R)}}, \tag{1.61}$$

where

$$R \equiv \frac{3\rho_b}{4\rho_\gamma} \propto \frac{\Omega_b h^2}{1+z}. \tag{1.62}$$

The oscillations end when the photons decouple and remove the radiation pressure. The stalled oscillations leave behind an imprint in the form of spherical shells of overdensities of a specific radius, set by Eq.(1.61). As the Universe evolves, dark matter (and subsequently baryonic matter) preferentially clumps on this scale (see Eq.(1.63) below), leaving behind a statistical signature in the correlation function of galaxies.

Selecting a galaxy as the centre of a sphere, there will be an increased probability of finding a galaxy (and hence a bump in the two-point correlation function) at scale s, where s is given by:

$$s = \int_{z_{\rm rec}}^{\infty} \frac{c_s \mathrm{d}z}{H(z)} = \frac{1}{\sqrt{\Omega_m H_0^2}} \frac{2c}{\sqrt{3 z_{\rm rec} R_{\rm eq}}} \ln\left[\frac{\sqrt{1+R_{\rm rec}} + \sqrt{R_{\rm rec}+R_{\rm eq}}}{1+\sqrt{R_{\rm eq}}}\right], \tag{1.63}$$

with $z_{\rm eq} = \Omega_m/\Omega_R - 1$ is the redshift of matter-radiation equality and "rec" refers to recombination. By using the value of Ω_b as determined by CMB observations (and assuming no exotic forms of radiation which would change the value of $z_{\rm eq}$), we find that $s \approx 150$ Mpc.

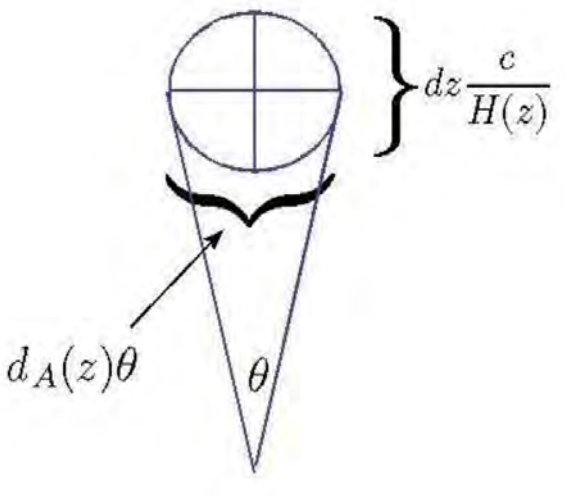


FIGURE 1.17: Illustration of a calibrated or absolute Alcock-Paczynski test, used to determine $d_A(z)$ and $H(z)$. If one assumes the object of interest is spherical in nature and of known radius, measurements of dz (the change in redshift, relating to the transverse size of the object) and θ (the angular size) determine $d_A(z)$ and $H(z)$ separately. If the diameter of the object is unknown, this reduces to an Alcock-Paczynski test, with only the assumption of sphericity, and only the product $d_A(z)H(z)$ can be accurately determined. [20]

From Eq.(1.33) and Eq.(1.36), we know that distance measures provide powerful probes of cosmology, including models of dynamical dark energy if the assumption of a cosmological constant is relaxed. Baryon Acoustic Oscillations have the power to provide both $H(z)$, through radial measurements, and $d_A(z)$, through transverse measurements. This method, known as a calibrated or absolute Alcock-Paczynski test, is illustrated in Figure 1.17 [20]. This can be applied in a statistical sense using measurements of the BAO. Detailed observations of the BAO using galaxy surveys provide a characteristic scale along the line of sight, $s_{\parallel}(z)$, and a scale tangential to the line of sight, $s_{\perp}(z)$, which can provide measurements of the desired cosmological functions by:

$$H(z) = \frac{c\Delta z}{s_{\parallel}(z)}, \tag{1.64}$$

and

$$d_A(z) = \frac{s_{\perp}}{\Delta\theta(1+z)}. \tag{1.65}$$

One of the main problems with BAO measurements is that they are model dependent. One must use a fiducial cosmological model to convert the redshifts to distance before obtaining the correlation function. This dependence is fairly weak, however, as long as the true cosmology is not far from the fiducial model [20].

The BAO was first detected as a bump in the two-point correlation function of the SDSS[14] galaxy catalogue [57]. This provided the first geometric evidence for dark energy, completely independent of supernovae observations (see Section 1.7.1). The BAO continues to be a powerful probe of large scale structure and cosmology. Figure 1.18 shows the recent BAO detection from BOSS[15], providing tight constraints on the cosmological parameters.

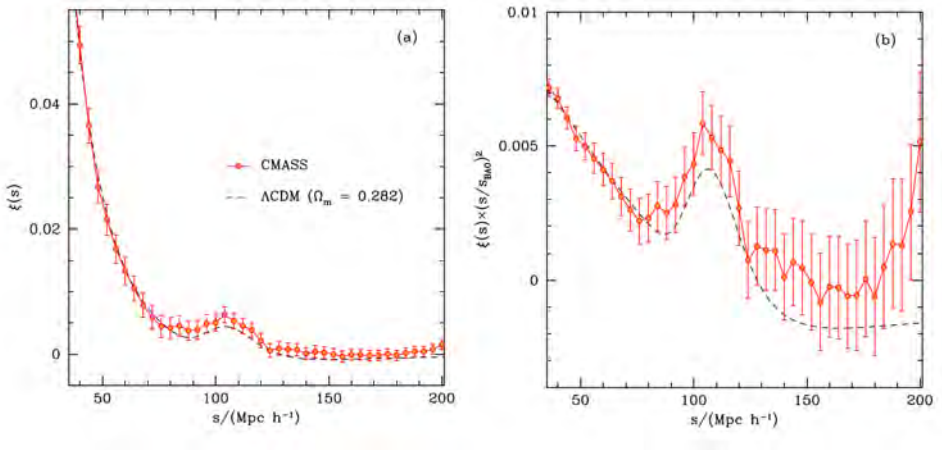


FIGURE 1.18: The spherically averaged two-point correlation function of the CMASS sample of data from the SDSS-III BOSS survey. Note that the y-axis in the second panel is multiplied by $(s/s_{\rm BAO})^2$, where $s_{\rm BAO} = 153.2$ Mpc (or 107.2 Mpc h^{-1}), to accentuate the BAO feature. The dashed line is the best fit ΛCDM model when combining the CMASS dataset with CMB data. It should be noted that the errors are highly correlated, making the model look like a worse fit than it is [58].

With its large HI survey [59], the SKA will provide tight constraints on the BAO in neutral hydrogen. As a recent example, Figure 1.19 shows the HI power spectrum (compare with Figure 1.16) from an SKA-like HI survey where intensity mapping of the HI is used (intensity mapping is simply where one uses the HI map without attempting to resolve it into individual galaxies first) [60]. Figure 1.20 shows the BAO wiggles from a recent SKA simulation and indicates the effects of foreground sources on the HI BAO signal [61].

[14] Sloan Digital Sky Survey, http://www.sdss.org
[15] SDSS-III Baryon Oscillation Spectroscopic Survey, http://www.sdss3.org/surveys/boss.php

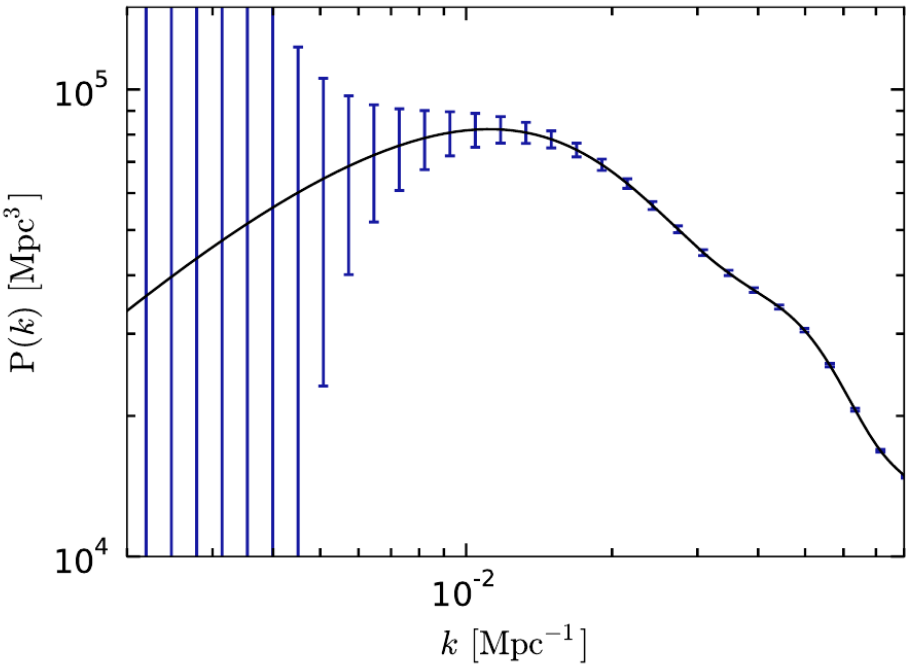


FIGURE 1.19: Simulated HI power spectrum with forecast constraints from an SKA-like survey [60].

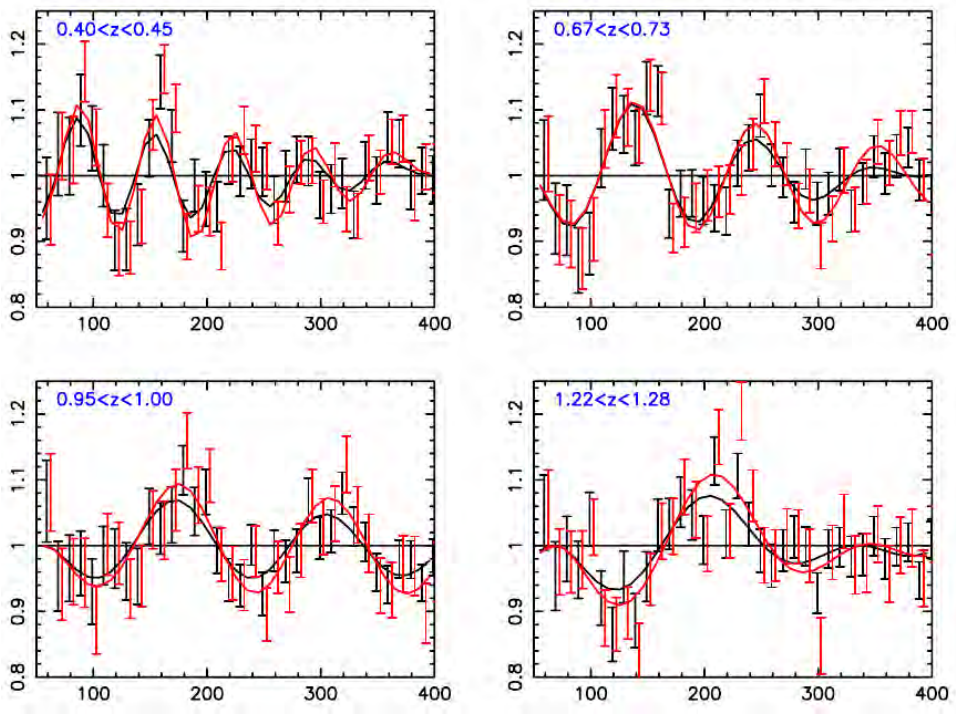


FIGURE 1.20: Simulated BAO wiggles for the SKA in example redshift slices. Black and red points are with the addition and removal of foregrounds respectively and the lines indicate best fit models [61].

1.8 Sources of Radio Frequency Radiation

Because Chapter 5 focuses on a technique for analysing radio astronomy observations, we introduce here an overview of some of the radio sources commonly observed in the Universe, data from which could be analysed with said technique.

One of the most common sources of radio frequency radiation are HII regions. HII is notation for ionized hydrogen and it produces weak radio signals by bremsstrahlung radiation, literally meaning "braking radiation". This is the radiation produced when a charged particle or ion is accelerated or decelerated by an electrostatic or magnetic force (in which case it is called magnetobremsstrahlung radiation). This type of radiation in an ionized cloud (such as an HII region) is known as free-free radiation because the electrons being accelerated usually have too much energy to be captured by the ions [62].

HII regions are formed in a sphere around an ionizing star, which emits UV photons of high enough energy to ionize hydrogen. These stars are either young, massive, hot stars or young, white dwarfs (remnants of dead stars) which are hot enough to emit in the UV [39]. The HII regions around white dwarfs form planetary nebulae, a misnamed but popularly imaged astronomical object [62].

Bremsstrahlung radiation has slightly different characteristics depending on the speed of the electrons. Gyro radiation is produced by electrons whose speeds are much slower than the speed of light, mildly relativistic electrons (whose kinetic energy is comparable to their rest mass) produce cyclotron radiation while sychrotron radiation is produced by highly relativistic electrons. Synchrotron radiation dominates the radio emission of active galactic nuclei (AGN) and also forms the low frequency ($\nu < 30$ GHz) component of normal galaxy emission [62].

Inverse-Compton scattering is a dominate source of free-free radio emission at high frequencies [62]. This occurs when a low frequency photon scatters off of an electron "stealing" some of the electron's energy and increasing its frequency. This produces the characteristic spectrum illustrated in Figure 1.21. A source which is emitting synchrotron radiation will sometimes produce Inverse-Compton radiation from the same relativistic electrons which produced the synchrotron radiation. This is known as synchrotron self-Compton radiation.

Figure 1.22 shows a schematic view of the popular unified model of active galactic nuclei (AGN) [39, 63]. AGN are, to quote Eric Wilcots, "galaxies behaving badly". These are the nuclei of

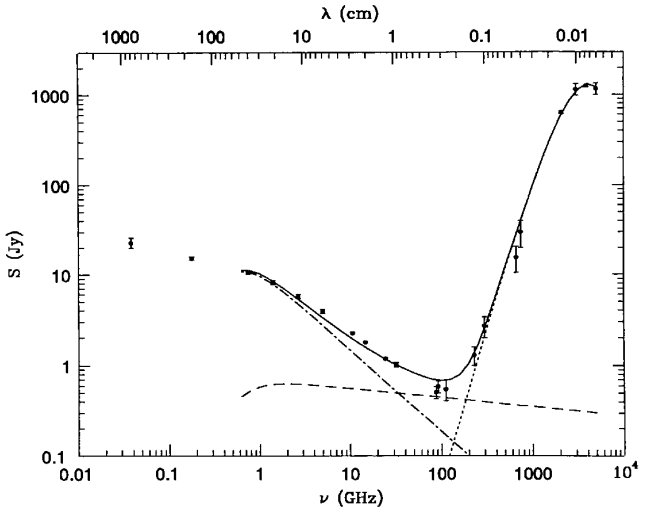


FIGURE 1.21: Flux density from a nearby starburst galaxy, M82. Synchrotron emission (dot-dashed line) from cosmic ray electrons and hot, young stars dominates at frequencies $\nu < 30$ GHz. Thermal emission (dashed line) from HII regions ionized by massive stars dominates at $30 < \nu < 200$ GHz. At frequencies about 200GHz, Inverse-Compton scattering begins to dominate as low frequency photons are scattered to higher frequencies. At frequencies below 1GHz, free-free absorption flattens out the whole spectrum [62].

active galaxies which have increased luminosity in at least some part of the electromagnetic spectrum [21]. They are thought to be powered by accretion onto a supermassive black hole, as shown in Figure 1.22 [63]. BLRG and NLRG refer to the broad line region and narrow line region galaxies respectively, referring to the types of emission line visible which depends on the viewing angle and the amount of obscuring dust. QSO refers to quasi-stellar object: an object whose AGN emission dominates normal galaxy emission. Otherwise the galaxy is known as a Seyfert I galaxy. Other types of AGN include Seyfert II galaxies, BL Lac objects and a distinction between radio loud and radio quiet AGN [62].

Many extragalactic sources emit powerful jets of material which radiate at radio frequencies from their central black hole. These jets can be classified into two categories: jets that appear to fade away far from the centre and jets that have edge-brightened lobes. These are named FR I and FR II sources respectively, after Fanaroff and Riley [64]. FR I sources tend to be fainter, with the cutoff being around $L_\nu \sim 10^{24}$W Hz^{-1} at 1.4 GHz [62].

Normal galaxies' radio emission is not powered by an AGN. It is dominated by a combination of free-free emission from HII regions around massive stars and synchrotron radiation from cosmic ray electrons, mostly from supernova remnants [21]. Because both these sources are not main sequence stars but are instead related to young stars and recent supernovae, radio continuum emission from normal galaxies is a tracer of recent star formation, uncontaminated by older

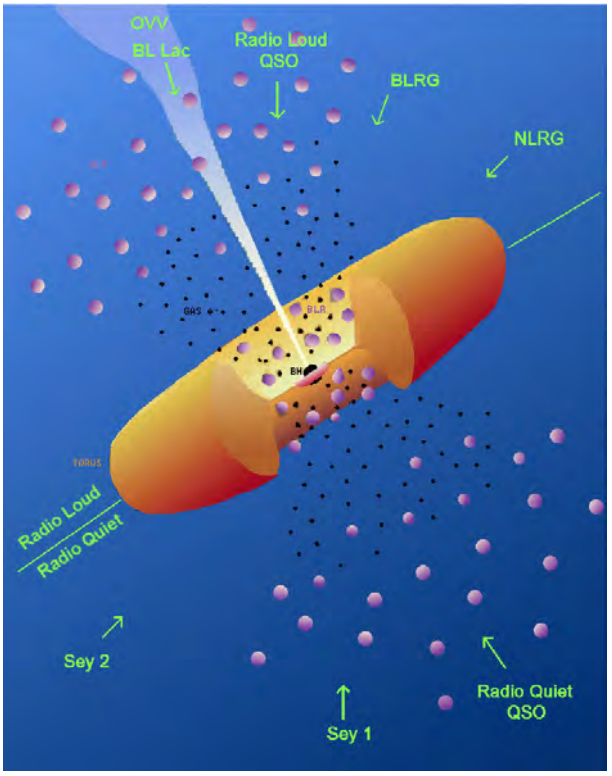


FIGURE 1.22: A cartoon drawing of a unified model of active galactic nuclei, where the different observed types of AGN are simply a function of angle of observation [63].

stars. This is especially true of starburst galaxies, which occur when galaxy-galaxy collisions trigger such intense star formation that the interstellar medium is depleted on time scales much shorter than 10^{10} years [62].

Pulsars are another source of radio emission which will be a focus area for the SKA [59]. Although the word pulsar is an amalgamation of the words "pulse" and "star", it merely appears to pulse but actually emits radiation continuously. A pulsar is a neutron star, the dense remnant of an exploded star [39]. The powerful magnetic field of the pulsar causes synchrotron radiation to be emitted in a narrow beam from the poles of the star. As the object rotates, it appears to pulse when the beam enters and exits our line of sight [62].

The extreme gravitational and magnetic environments of pulsars make them excellent astronomical laboratories [62]. Indeed, one of the SKA key science projects is using pulsars to test theories of gravity. By studying thousands of pulsars, it is likely the SKA will locate a pulsar in orbit around a black hole providing the first tests of General Relativity in the strong field limit [65]. Additionally, the thousands of pulsars monitored by the SKA will form a pulsar timing array network with which gravitational waves may be detected and studied [59, 65].

Chapter 2

An Introduction to Radio Astronomy

2.1 Introduction

Radio astronomy is the study of the Universe in the radio regime of the electromagnetic spectrum, 10MHz to 1THz. This window of observation is huge, spanning roughly five decades. Although this means that a wide variety of astronomical sources and types of radiation are visible in the radio, it also means that an equally wide variety of techniques and technologies need to be employed to study the entire radio regime [62].

Radio astronomy is unique not only in the astronomical sources visible but also the relative ease with which they can be studied: the atmosphere is transparent to large portions of the radio spectrum (see Figure 2.1) [66], the sun is a faint radio source (in many frequency bands) allowing radio observations to be done day or night and scattering of certain radio waves off of interstellar dust is almost negligible [67], meaning regions obscured in the optical, such as the centre of our galaxy, are clearly visible in certain radio wavelength bands [62].

Radio astronomy began with Karl Guthe Jansky when he was working for Bell Telephone Laboratories [67]. He was instructed by his employer to find the source of natural radio static which could interfere with their new telephone service based on shortwave radio waves. Jansky, using his purpose built antenna, discovered a "hiss" coming from galactic centre. He published his results in 1933 [68] but, as Bell Laboratories had no use for further research, he was reassigned

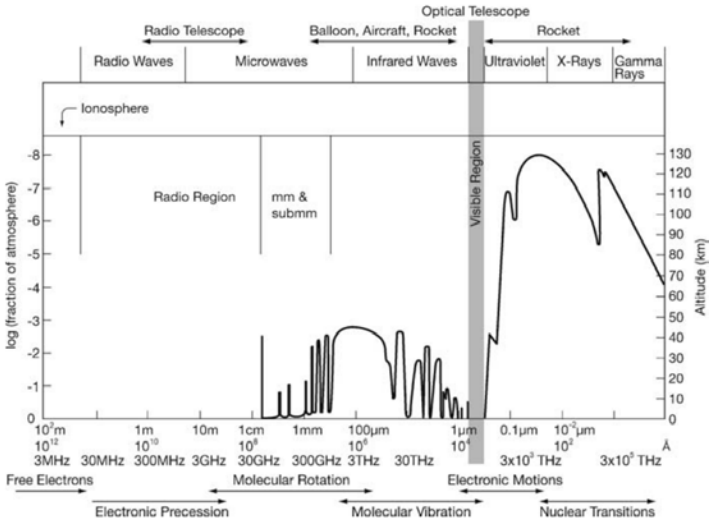


FIGURE 2.1: The transmission of the Earth's atmosphere for electromagnetic radiation at the height at which the radiation is attenuated by a factor of 1/2 [66].

to a different task. However, Jansky is honoured by having the unit of measure for radio flux named after him. One Jansky is tiny, 1 Jy $= 10^{-26}\,\mathrm{W\,m^{-2}\,Hz^{-1}}$, a testament to how faint most radio sources are [62].

Although at the time most astronomers were skeptical of Jansky's findings, he inspired amateur radio operator and professional radio engineer Grote Reber to build the world's first parabolic radio antenna, at his own expense, who eventually managed to successfully map the galaxy at a frequency of 160MHz thereby confirming Jansky's discovery [67]. Although the Second World War intervened with astronomical research, it stimulated great progress in radio and radar technology and the field of radio astronomy grew from there [62].

2.2 Fundamentals of Radio Observations

Because the frequency range of radio astronomy is so large, several different types of radio telescopes exist, from the simple dipole antenna to the more typical parabolic reflector [66]. A reflector antenna usually consists of a paraboloid primary mirror which reflects and focuses radio waves back onto a feed antenna, such as a waveguide horn or dipole, connected to a receiver. Reflector antennae work over a large range of frequencies and changing the operating frequency only requires changing the receiver at the focal point [62].

From antenna theory, the average effective collecting area of any lossless antenna, regardless of shape, is [62, 66]:

$$\langle A_e \rangle = \frac{\lambda^2}{4\pi}, \tag{2.1}$$

where λ is the wavelength. This means that any non-directional antenna operating at short wavelengths will have a very small collecting area. Consequently, arrays of dipole antennae are only useful for low frequency observations.

Central to understanding a radio telescope is the concept of a power pattern or beam. A telescope antenna is often treated as a transmitter instead of a receiver for convenience, since the properties are almost identical (reciprocity theorem) [66]. Then the power pattern is the angular distribution of radiated power, often normalised to unity at the peak. For example, the power pattern of a simple dipole antenna is

$$P \propto \sin^2\theta. \tag{2.2}$$

When reversed, this can be thought of as the angular pattern within which the antenna will receive radiation.

The beam strongly influences the final image obtained. Typically, the beam pattern for a standard reflecting telescope consists of a main beam and sidelobes. The main beam is defined as the region containing the principal response out to the first zero [62] and the sidelobes are responses outside this region. Telescope designers will usually try to minimise the sidelobes as much as possible to reduce contaminating radiation from bright sources far away from the target or even from the warm ground that can leak into the sidelobes.

To illustrate the main beam and sidelobes, consider a simple, one dimensional uniformly illuminated aperture. In this case, the beam pattern is:

$$P(\theta) \propto \mathrm{sinc}^2\frac{\theta D}{\lambda}, \tag{2.3}$$

where D is the diameter of the aperture. This is shown in Figure 2.2. The beam pattern is given by the Fourier transform of the antenna aperture and this is true for any antenna.

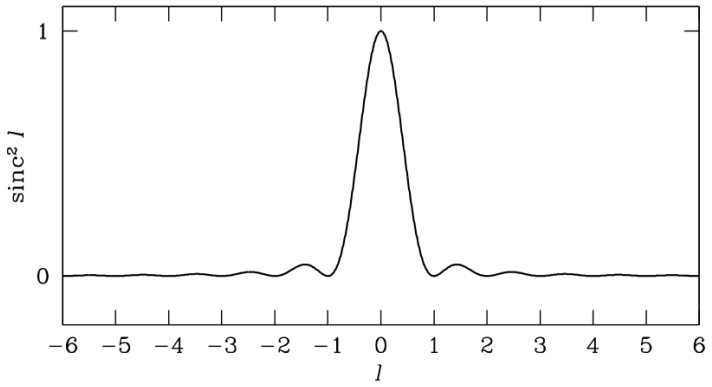


FIGURE 2.2: The beam pattern of a uniformly illuminated unit ($D/\lambda = 1$) aperture. For large ($D >> \lambda$) apertures, the zeros at $l = 1°, 2°, ...$ appear at the angles $\theta = \lambda/D°, 2\lambda/D°, ...$ [62] The main beam and smaller sidelobes are clearly visible.

The half-power beam width (HPBW) or the full width at half maximum (FWHM) is defined as the angle between the half-power points of the beam. For most single dish radio telescopes [67]:

$$\theta_{\text{HPBW}} \approx 1.2\frac{\lambda}{D}. \tag{2.4}$$

The HPBW is sometimes called the resolving power of the telescope because two point sources separated by the HPBW can just be resolved by the Rayleigh criterion when the maximum response of one source coincides with the minimum response of the other [62].

The antenna temperature is defined as [67]:

$$T_A = \frac{P_\nu}{k}, \tag{2.5}$$

where P_ν is the power per unit frequency and k is the Boltzman constant. The antenna temperature is a very practical unit to compare sources to because it can be calibrated against hot and cold "loads" (or matched resistors) connected to the receiver input. The antenna temperature can be directly compared with the brightness temperature of a source, defined (for any I_ν) as

$$T_b(\nu) = \frac{I_\nu c^2}{2k\nu^2}, \tag{2.6}$$

where I_ν is the intensity of the source. The brightness temperature is only equal to the physical kinetic temperature for opaque sources [62].

2.3 Interferometry

The largest steerable radio telescopes have diameters of about 100 m, which is not large enough to achieve sub-arcsecond resolution at radio wavelengths [62]. Eq.(2.4) implies the resolution of a 100 m telescope at the high frequency of 86GHz is only about 9 arcseconds. Interferometers, telescopes made up of two or more dishes, have revolutionised the field of precision radio astronomy, allowing extremely high resolution imaging and photometry [66].

2.3.1 The Two-Element Interferometer

The simplest interferometer is made up of two radio telescopes whose voltage outputs are correlated (multiplied and averaged). Even complex interferometers of N elements can simply be treated as $N(N-1)/2$ independent pairs so we begin with the two-element narrowband interferometer [62].

Figure 2.3 shows two antennae, separated by a baseline vector of length b, both viewing the same source in the direction $\hat{s}$. Light from the source must travel an extra distance $\vec{b}\cdot\hat{s} = b\cos(\theta)$ to reach antenna 1, resulting in a geometric delay, $\tau_g = \vec{b}\cdot\hat{s}/c$ between the two signals. If the interferometer is quasi-monochromatic (only responds to a narrow frequency band), then the output voltages of the two antennae are

$$V_1 = V\cos[\omega(t-\tau_g)] \quad \text{and} \quad V_2 = V\cos(\omega t), \tag{2.7}$$

where t is time and $\omega = 2\pi\nu$ is angular frequency.

The correlator first multiplies the two voltages (the cross in Figure 2.3)

$$V_1V_2 = V^2\cos(\omega t)\cos[\omega(t-\tau_g)] = \left(\frac{V^2}{2}\right)[\cos(2\omega t - \omega\tau_g) + \cos(\omega\tau_g)], \tag{2.8}$$

and then takes the time average (the $\langle\rangle$ symbol in Figure 2.3) long enough to remove the high frequency term $\cos(2\omega t - \omega\tau_g)$ from the final output R [62]:

$$R = \langle V_1V_2\rangle = \left(\frac{V^2}{2}\right)\cos(\omega\tau_g). \tag{2.9}$$

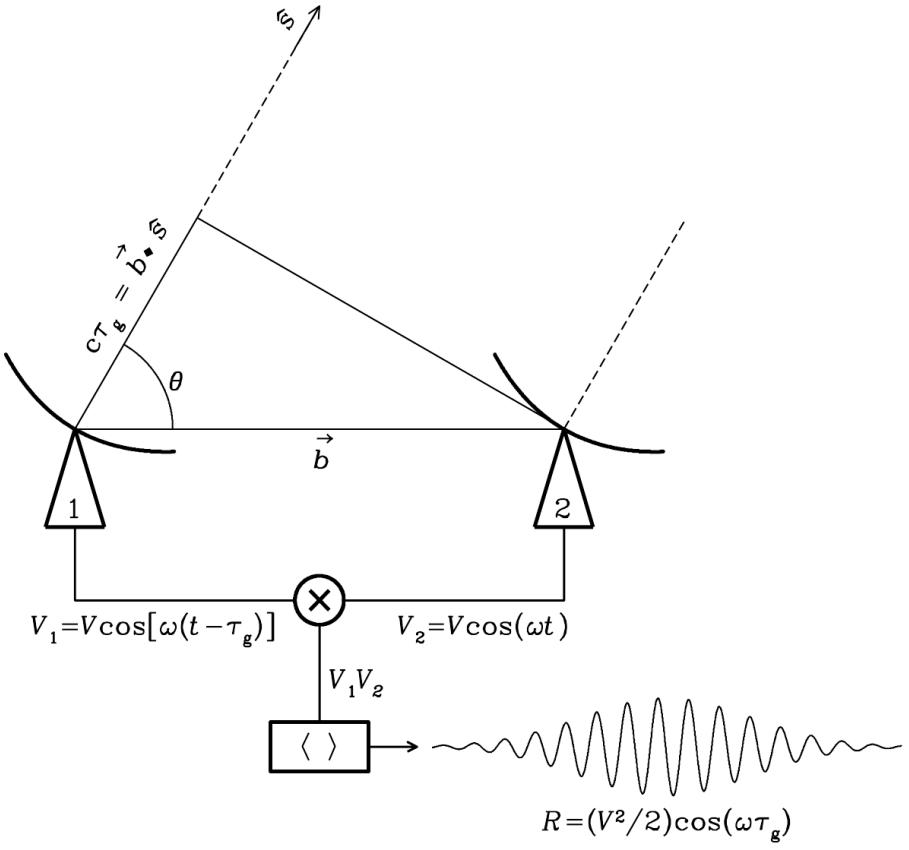


FIGURE 2.3: Two-element narrowband interferometer, observing in a narrow frequency range centred on $\nu = \omega/(2\pi)$. V_1 and V_2 are the time-dependent output voltages of antenna 1 and 2. $\hat{s}$ is the unit vector in the direction of a distant point source and $\vec{b}$ is the vector baseline between the antennae. The output voltage V_1 is retarded by the geometric delay $\tau_g = \vec{b} \cdot \hat{s}/c$. V_1 and V_2 are multiplied and time averaged by the correlator. R, the amplitude of the output response, is proportional to the source flux density, whereas the phase depends on the delay and the frequency. The sinusoidal fringes occur if the source direction in the interferometer frame is changing at a constant rate $d\theta/dt$. The broad Gaussian envelope of the fringes is caused by attenuation in the primary-beam if the individual dishes do not track the source [62].

Another advantage of interferometry is a dramatic reduction in noise [66]. Any uncorrelated noise in the receivers does not appear in the final output. Thus interferometers are far less sensitive to receiver noise and atmospheric disturbances than single dishes [62].

The correlator output voltage varies sinusoidally with any change in source direction in the frame of the interferometer. These are called fringes and have a phase of

$$\phi = \omega\tau_g = \frac{\omega}{c} b \cos(\theta), \tag{2.10}$$

where θ is the angle between the source and the baseline, as shown in Figure 2.3. Thus the fringe phase depends on θ as follows:

$$\frac{d\phi}{d\theta} = \frac{\omega}{c} b \sin(\theta) = 2\pi \frac{b \sin(\theta)}{\lambda}. \tag{2.11}$$

The fringe phase is an highly sensitive measurement of the source position if the projected baseline, $b \sin(\theta)$, is many wavelengths long. Interferometers can measure source position with superior accuracy to that of single dish telescopes because fringe phase is not affected by small tracking errors (depending on time not angles) and is also unaffected by the plane-parallel component of atmospheric effects, since it influences both interferometer elements equally [62].

If the individual antennae of the two-element interferometer were isotropic, the response of the interferometer to a point source would be a sinusoid spanning the sky, sensitive only to the angular period $\lambda/(b \sin\theta)$ [62]. Since the antennae are directive, the response is actually this sinusoid multiplied by the power pattern of the individual antennae and is called the primary beam [66]. This beam is usually Gaussian, as illustrated in Figure 2.3. The primary beam of two directive antennae responds to a finite range of angular frequencies, centred on $b\sin\theta/\lambda$. The angular frequency response cannot extend to zero since the antenna diameter must always be smaller than the baseline (or the dishes would overlap) and so a directional antenna cannot detect an isotropic source (for example the CMB) [62].

To improve the point source response of an interferometer, more Fourier modes and hence more baselines must be added. The synthesised beam, that is the point source response obtained by averaging the primary beams of individual pairs of antennae, approaches Gaussian as more antennae are added to the interferometer. For a stationary source (a source whose brightness distribution does not change appreciably over time), a two-element moveable interferometer can make $N(N-1)/2$ observations to duplicate one observation with an N-element interferometer. See Section 2.3.4 for an illustration of how adding more baselines improves the synthesised beam of the telescope [62].

The synthesised beam is also known as the point spread function of the telescope, since it is exactly the telescope's response to a single point source. Figure 2.4 shows the synthesised beam of WSRT[1] as an example.

[1] Westerbork Synthesis Radio Telescope, https://www.astron.nl/radio-observatory/astronomers/wsrt-astronomers

FIGURE 2.4: The WSRT synthesised beam, also known as the point spread function. Note the strong side lobes around the main beam [69].

2.3.2 Extended Emission

Many radio sources have extended emission. We can treat this extended source, with sky brightness distribution $I_\nu(\hat{s})$, as a sum of independent point sources, to obtain the response [62]:

$$R_c = \int I_\nu(\hat{s})\cos(2\pi\nu\vec{b}\cdot\hat{s}/c)d\Omega = \int I_\nu(\hat{s})\cos(2\pi\vec{b}\cdot\hat{s}/\lambda)d\Omega, \tag{2.12}$$

where the integral is performed over solid angle $d\Omega$ (see Eq.(1.14)).

The problem is, this response function is only sensitive to the even part of the source brightness distribution. To detect the odd component, we need a "sine" correlator with an odd output, $R_s = (V^2/2)\sin(\omega\tau_g)$. This can be implemented using a second correlator which inserts a 90° phase delay into the output of one antenna. Thus the output of this correlator for an extended source is [62]:

$$R_s = \int I_\nu(\hat{s})\sin(2\pi\vec{b}\cdot\hat{s}/\lambda)d\Omega. \tag{2.13}$$

Combining the "cosine" and "sine" correlators gives us a "complex" correlator, so called because it is convenient to write the sine and cosine components as complex exponentials with the identity:

$$e^{i\phi} = \cos(\phi) + i\sin(\phi). \tag{2.14}$$

Now we define the complex visibility as $V = R_c - i\,R_s$ and write it as:

$$V = Ae^{-i\phi}, \tag{2.15}$$

where

$$A = (R_c^2 + R_s^2)^{1/2} \tag{2.16}$$

is the visibility amplitude, and

$$\phi = \arctan(R_s/R_c) \tag{2.17}$$

is the phase.

Thus, we can combine the two sine and cosine responses to obtain the complex visibility for an extended source:

$$V_\nu = \int I_\nu(\hat{s})\exp(-i2\pi\vec{b}\cdot\hat{s}/\lambda)\,d\Omega. \tag{2.18}$$

2.3.3 Bandwidth and Time Smearing

Equation (2.18) can be generalised to include a finite bandwidth of $\Delta\nu$ centred on a frequency of ν_c [62]:

$$V = \int \left[(\Delta\nu)^{-1}\int_{\nu_c-\Delta\nu/2}^{\nu_c+\Delta\nu/2} I_\nu(\hat{s})\exp(-i2\pi\vec{b}\cdot\hat{s}/\lambda)d\nu\right] d\Omega. \tag{2.19}$$

$$= \int \left[(\Delta\nu)^{-1}\int_{\nu_c-\Delta\nu/2}^{\nu_c+\Delta\nu/2} I_\nu(\hat{s})\exp(-i2\pi\nu\tau_g)d\nu\right] d\Omega. \tag{2.20}$$

The integral in the square brackets is the Fourier transform of a rectangle function. Thus:

$$V = \int I_\nu(\hat{s})\mathrm{sinc}(\Delta\nu\tau_g)\exp(-i2\pi\nu\tau_g)\,d\Omega, \tag{2.21}$$

where the reader is reminded that $\mathrm{sinc}(x) = \sin(x)/x$. So the fringe amplitude is attenuated by a factor of $\mathrm{sinc}(\Delta\nu\tau_g)$. This can be eliminated in any one direction, $\hat{s_0}$, called the delay centre, by introducing a compensating delay of $\tau_0 \approx \tau_g$ in the signal path of antenna 2. This must be constantly adjusted by digital electronics as the Earth turns [62].

Since the geometric delay varies with direction, this compensation can only be applied in one direction at a time. This defines a maximum field of view which is dependent on bandwidth. This results in the requirement of [62]

$$\Delta\theta\Delta\nu \ll \theta_s\nu, \tag{2.22}$$

where $\theta_s \approx \lambda/(b\sin\theta)$. Larger $\Delta\theta$ will result in bandwidth smearing: radially broadening the synthesised beam [67].

Satisfactory wide field images can only be made with large bandwidths if the bandwidth is divided into narrow frequency channels. Time smearing causes a similar problem [67]. If an observation is carried out for too long, such that the Earth's rotation causes the source to move appreciably while the correlator is time averaging, this will cause additional broadening of the synthesised beam [67]. A similar relation holds for the length of time of an observation [62]:

$$\Delta\theta\Delta t \ll \frac{\theta_s P}{2\pi}, \tag{2.23}$$

where $P \approx 86164\,s$ is the Earth's sidereal rotation period.

2.3.4 Earth-Rotation Aperture Synthesis

The rotation of the Earth can be used as a way of increasing the baseline coverage of an interferometer. This is particularly useful for an interferometer such as the Westerbork Synthesis Radio Telescope[2], which is arranged in an East-West line. This confines the baselines to a single two dimensional plane, which has the computational advantage that the brightness distribution of a source, $I_\nu(\hat{s})$, is simply the Fourier transform of the measured visibilities [62].

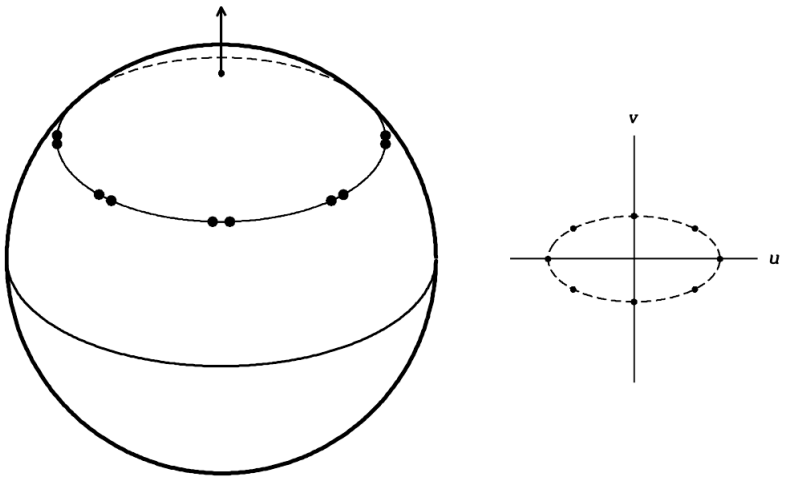


FIGURE 2.5: An example of Earth-rotation aperture synthesis with an east-west two-element interferometer, plotted at three hour intervals. On the right, the projected baseline is shown in the uv-plane during the same 12-hour period [62].

2.3.4.1 The uv-plane

Figure 2.5 illustrates Earth-rotation aperture synthesis. We define u as the East-West component of the projected baseline, measured in wavelengths, and v as the North-South component. Figure 2.5 also shows the ellipse traced out in the so-called uv-plane by the projected baseline as the Earth rotates. If more baselines are added (or if the telescopes are moved to a new configuration), the uv coverage becomes a number of concentric ellipses having the same shape [62]. Figure 2.6 illustrates how more baselines and Earth-rotation aperture synthesis fills in the uv-plane and improves the shape of the synthesised beam.

2.3.5 Interferometers in Three Dimensions

For interferometers such as the Very Large Array[3] not arranged in an east-west line, the Earth's rotation causes their baselines to fill a three dimensional volume instead of lying on a plane, although on short timescales these telescopes can take two dimensional "snapshot" observations. In the case of three dimensions, we add an extra coordinate, w, to form the (u, v, w) coordinate system to describe any baseline vector $\vec{b}$ [62]. The w axis is usually chosen such that it contains the source and u and v are at right angles to it pointing east and north respectively [66]. An arbitrary unit vector, $\hat{s}$, has components (l, m, n) which are the projections of $\hat{s}$ onto the coordinate axes and are known as direction cosines [62].

[2]Westerbork Synthesis Radio Telescope, http://www.astron.nl/radio-observatory/public/public-0
[3]Very Large Array, http://www.vla.nrao.edu/

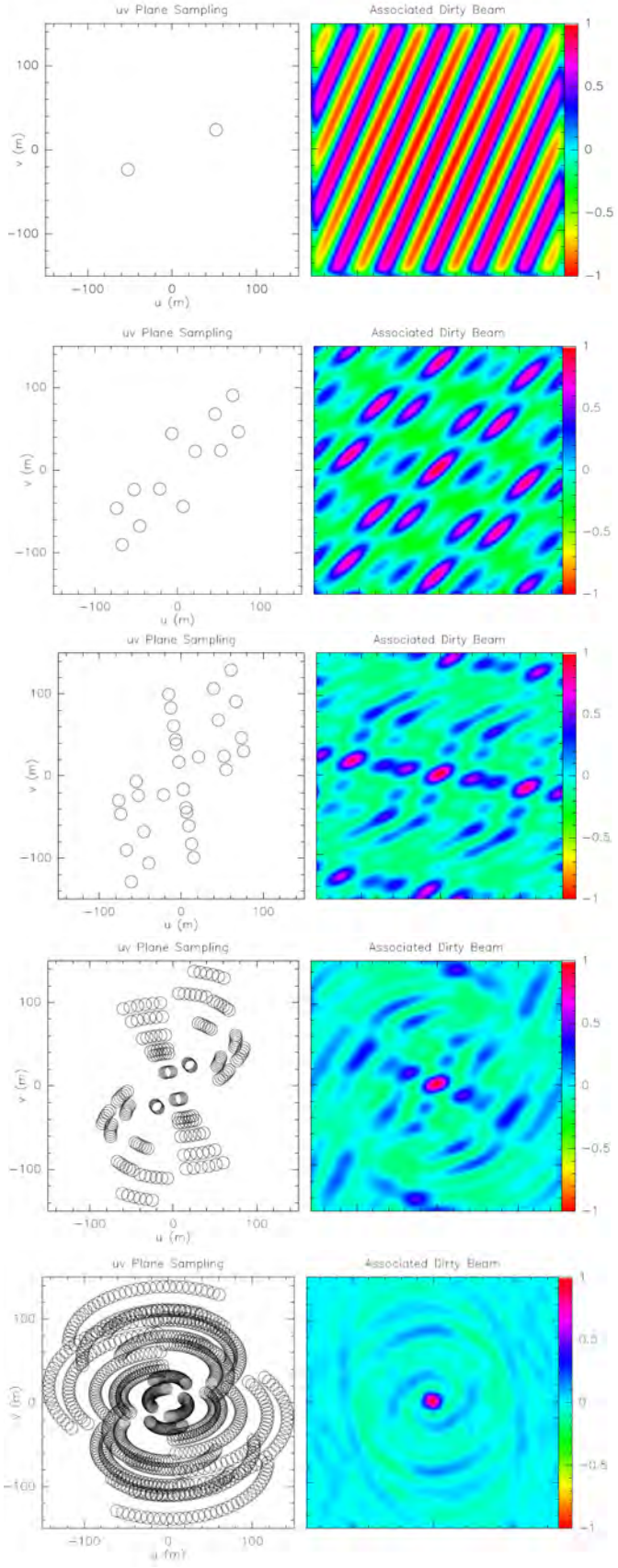


FIGURE 2.6: Illustration of aperture synthesis. *Left column: uv*-plane. *Right column:* Synthesised beam. In the first three rows, 2, 4 and 6 antennae are added to the array. In the last two rows, Earth rotation aperture synthesis is applied to subsequently longer observations [70].

Eq.(2.18) can be generalised to three dimensions (ignoring polarisation and instrumentation effects and assuming propagation of light through empty space) as [62]:

$$V_\nu(u,v,w) = \int\int \frac{I_\nu(l,m)}{(1-l^2-m^2)^{1/2}} \exp[-i2\pi(ul+vm+wn)]\, dldm, \tag{2.24}$$

which is *not* a three dimensional Fourier transform, making it computationally challenging to compute.

However, Eq.(2.24) can be converted into a form which is a Fourier transform by making a few assumptions and restrictions on observations. First we will only consider directions close to the w-axis (close to the source) such that $n = cos\theta \approx 1 - \theta^2/2$, where θ is the angle between $\hat{s}$ and the w-axis. Then [62]

$$V_\nu(u,v,w) \approx \exp(-i2\pi w) \int\int \frac{I_\nu(l,m)}{(1-l^2-m^2)^{1/2}} \exp[-i2\pi(ul+vm-w\theta^2/2)]\, dldm. \tag{2.25}$$

We force the $\exp(i2\pi w\theta^2/2)$ factor close to unity by requiring that $w\theta^2 \ll 1$. That means imaging a field small enough such that $\theta \ll w^{-1/2} \approx (\lambda/b)^{1/2}$. Then:

$$V_\nu(u,v,w) \approx \exp(-i2\pi w) \int\int \frac{I_\nu(l,m)}{(1-l^2-m^2)^{1/2}} \exp[-i2\pi(ul+vm)]\, dldm. \tag{2.26}$$

Larger fields can be imaged by breaking it up into mosaics and then merging them together to make the final image [62].

2.3.6 Sensitivity

Sensitivity is described as the ability of a telescope to detect faint sources [67]. The point source sensitivity of an N element interferometer is [62]:

$$\sigma_S = \frac{2kT_{\text{sys}}}{A_{\text{eff}}[N(N-1)\Delta\nu_{\text{RF}}\tau]^{1/2}}, \tag{2.27}$$

where T_{sys} is the system temperature of the antenna, A_{eff} is the effective collecting area of a single antenna, $\Delta\nu_{\text{RF}}$ is the band width of the interferometer and τ is the length of time of the observation. The sensitivity gets worse as the system temperature increases but improves as the number, area, or band width of antennae is increased, or or longer observations are taken [62].

Although the point source sensitivity of an interferometer is comparable with that of a single dish antenna of the same total area, the brightness sensitivity is much worse because the synthesised beam solid angle is much smaller for an interferometer (owing to the gaps between telescopes). A high resolution interferometer is unable to detect sources of low surface brightness, no matter how bright the total flux is [62].

2.4 Introducing the RIME

There are many effects that can interfere with a radio signal in the path between the source and the final output of the telescope. There may be mispointings in the antennae, complicated beam effects, ionospheric and atmospheric effects, to name a few [71]. The Radio Interferometry Measurement Equation (RIME) [72, 73] provides a simple, mathematically rigorous way to combine these effects to model the output from an initial sky model.

The RIME in its simplest form, is written as [71]:

$$\mathsf{V}_{pq} = \boldsymbol{J}_p \mathsf{B} \boldsymbol{J}_q^H, \tag{2.28}$$

where V_{pq} is the visibility matrix, B is the brightness matrix and $\boldsymbol{J}$ is the Jones matrix [74]. The brightness matrix contains the sky model while the Jones matrices for antennae p and q contain the instrumental, atmospheric and other intervening effects [71].

In Chapter 5, we will discuss the RIME in detail and use it in developing a new technique for analysing radio astronomy observations.

Chapter 3

Bayesian Statistics

3.1 Introduction

The use of Bayesian statistics in astronomy and cosmology has increased dramatically over the last decade, as demonstrated by Figure 3.1. Although the work of Reverend Thomas Bayes was published back in 1763 [75], it took the massive increase in available data, as well as computational power, over the last decade for Bayesian statistics to become a commonly used tool in the fields of astronomy and cosmology.

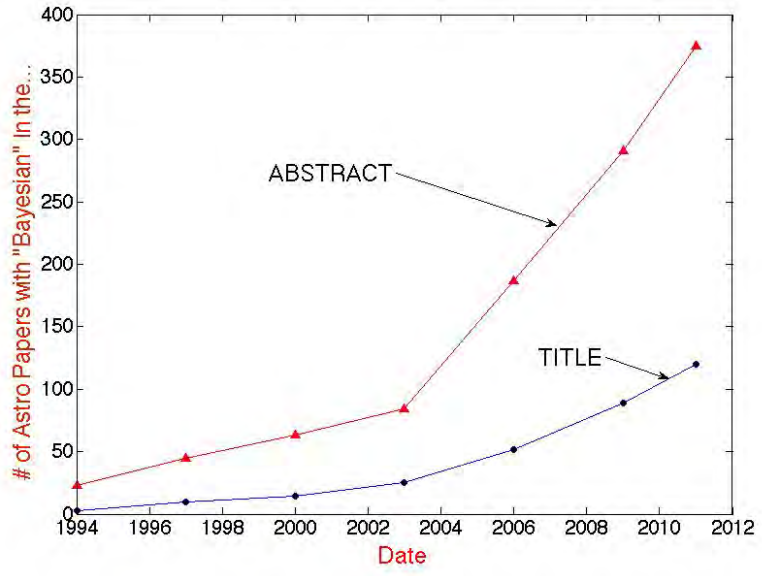


FIGURE 3.1: The increase of the use of Bayesian statistics in astronomy and cosmology papers as a function of time (Bruce Bassett - private communication).

The primary difference between a Bayesian and a (more traditionally used) frequentist point of view is the definition of probability. A frequentist defines probability in terms of trials: the probability is equal to the number of times an event occurs divided by the number of trials, in the limit of infinite repetitions, while a Bayesian defines probability as the degree of belief about a proposition [76]. The Bayesian definition of probability can be applied in real world problems where an infinite or even a large number of trials is impossible. For example, cosmic variance limits the possible number of samples on large scales, so statistical methods based on asymptotic limits for the data are not applicable.

Bayesian statistics recover frequentist results in the asymptotic limit, for particular cases where frequentist statistics are applicable at all. Also, nuisance parameters (for example, parameters to do with instrumental uncertainty or unknown underlying distributions) are easily incorporated in Bayesian inference. If nuisance parameters are ignored, or fixed to their best fit value, the uncertainty on the parameters of interest can be seriously underestimated. Lastly, prior information can and should be applied to a Bayesian inference problem, particularly in the case where the parameter space should be restricted due to physical and theoretical constraints [76].

3.2 Fundamentals of Bayesian Statistics

3.2.1 Probability Theory

There has been some question in the past as to whether a probability framework such as Bayesian statistics should be applied to inference problems. An interesting approach in [77] interprets probability theory as an extension of boolean logic to deal with propositions in the presence of uncertainties. [76] argues that with this interpretation, as well as the advantages mentioned above, the framework of Bayesian statistics is ideal for inference. With this view in mind, there are several rules probabilities follow which can be used in inference problems [78]:

Independence - If two random variables (or propositions), A and B are independent, then

$$P(A, B) = P(A)P(B). \tag{3.1}$$

This equation should be read as "The probability of A *and* B is equal to the probability of A times the probability of B". $P(A, B)$ is called the joint probability and is the probability for both A and B to be true.

Sum rule - If $\bar{A}$ denotes the negation of A and I denotes some information which is assumed to be true [76], then:

$$P(A|I) + P(\bar{A}|I) = 1. \tag{3.2}$$

$P(A|I)$ should be read as "The probability of A *given* I". We will continue to use I to explicitly indicate where assumptions may have been made. For example, in inferring something from data, we may be implicitly assuming that the data are an unbiased representation of reality. This assumption should enter under I.

Conditional probability -

$$P(A|B) = \frac{P(A,B)}{P(B)}, \tag{3.3}$$

where $P(B) \neq 0$.

Product rule - The product rule follows directly from the conditional probability:

$$P(A,B|I) = P(A|B,I)P(B|I) = P(B|A,I)P(A|I). \tag{3.4}$$

Marginal probability - Combining the sum and the product rule determines the marginal probability of one variable alone:

$$P(B|I) = \sum_A P(A,B|I). \tag{3.5}$$

3.2.2 Bayes' Theorem

Bayes' Theorem [75] is obtained directly from the product rule. Noting that $P(A,B|I) = P(B,A|I)$, we find [76]:

$$P(A|B,I) = \frac{P(B|A,I)P(A|I)}{P(B|I)}. \tag{3.6}$$

To use Bayes' Theorem for inference, we replace A by the hypothesis we want to test, H, and B by the data, D:

$$P(H|D,I) = \frac{P(D|H,I)P(H|I)}{P(D|I)}. \tag{3.7}$$

$P(H|D,I)$ is called the *posterior probability*, which represents our degree of belief in the hypothesis, given the data. On the right hand side, $P(D|H,I)$ is called the *likelihood function* and

encodes how the plausibility of the hypothesis changes in light of the new data. This quantity is usually fairly easily calculable in inference problems whereas directly evaluating the posterior is more challenging. Note that as the likelihood is a function of the hypothesis (and hence its parameters), it is not strictly a probability distribution. The likelihood is often simply written as $\mathcal{L}(H)$. $P(H|I)$ is called the *prior probability* and is obtained from any existing information one has about the hypothesis, usually in the form of physical limits on parameters or previous experimental constraints. $P(D|I)$ is called the *marginal likelihood* or the Bayesian evidence and is calculated by [76]:

$$P(D|I) = \sum_{H} P(D|H, I)P(H|I) \tag{3.8}$$

where H ranges over all possible outcomes for hypothesis H.

It should be noted that the sequence from prior to posterior is logical, not temporal, so the prior may well be specified after data has been taken. Thus Bayes' Theorem is a rule for updating current belief in an hypothesis based on new data.

3.2.3 Priors and Assumptions

The fact that priors are subjective and the Bayesian framework gives no guidance as to how to select them has been a subject of criticism for Bayesian statistics. However, a fundamental principle of Bayesian statistics is precisely that there cannot be inference without assumption [76] and the prior is a way of encoding those assumptions rather than ignoring them. Although two scientists may have two different priors for the same problem, as long as those priors are non-zero in regions where the likelihood is large, then repeated application of Bayes' Theorem will converge to the same, objective posterior for the problem. However, in cases where the data are not very constraining, the likelihood may be less informative than the prior and hence care should always be taken to determine the effect of the prior on the posterior. Commonly used priors include: uninformative or flat priors, where the prior is constant within a given parameter range and zero outside, Gaussian priors centred on a particular parameter value and the Jeffrey's prior, where $P(\theta) \propto \theta^{-1}$ where θ is the parameter of interest.

3.3 Parameter Inference with Bayesian Statistics

The first step in Bayesian inference is to choose a model, $\mathcal{M}$ whose parameters we wish to estimate given the data. Call this list of parameters θ. We then choose a prior for the parameters, which should include previous knowledge about the parameters. The next step is to define a likelihood function for the problem. Its form will usually reflect how the data were obtained. It is particularly common in cosmology to assume the noise is Gaussian and use a normal distribution for the likelihood but for data involving photon counts, for example, the likelihood would be described by a Poisson distribution. The final step is to compute the joint posterior:

$$P(\theta|D,\mathcal{M}) = \mathcal{L}(\theta)\frac{P(\theta|\mathcal{M})}{P(D|\mathcal{M})}. \tag{3.9}$$

In general, inference problems involve several parameters, which are often not independent. There may also be parameters called nuisance parameters, which we are not interested in, but which are unknown and may influence the parameters of interest and so cannot be ignored. In order to find the posterior probability for just one parameter (to obtain both the estimate of the parameter and the uncertainty on that estimate) or to remove the effect of nuisance parameters, we must compute the marginal probability. Let ϕ be the parameter(s) of interest and ψ be the nuisance/unwanted parameters. The marginal probability for ϕ is [76]:

$$P(\phi|D,I) = \sum_{\psi} P(\phi,\psi|D,I). \tag{3.10}$$

For most inference problems, the parameters are continuous and the sum becomes an integral. The full marginal posterior for a particular model $\mathcal{M}$ is [76]

$$P(\theta|D,\mathcal{M}) \propto \int \mathcal{L}(\phi,\psi)P(\phi,\psi|\mathcal{M})d\psi. \tag{3.11}$$

Very few problems have an analytical solution for the posterior. For most real-world problems, this multi-dimensional integral must be evaluated numerically. Fortunately, due to the increase in cheap computational power in recent years, as well as the advance of algorithms designed to solve this problem, Bayesian inference is now a practical and vital tool in many fields. One such algorithm, Markov Chain Monte Carlo, allows one to map out the posterior with very little increase in computational complexity as the number of parameters increases.

3.3.1 Markov Chain Monte Carlo Techniques

Markov Chain Monte Carlo (MCMC) techniques are a class of numerical algorithms designed to sample the posterior for a problem. Any MCMC algorithm will build up a sequence of points in parameter space, called a "chain", whose density is proportional to the posterior probability [76]. While many MCMC algorithms exist, we will focus on one of the simplest and most popular, the Metropolis-Hastings [79, 80] algorithm, to illustrate the use of MCMC methods in parameter estimation.

The Metropolis-Hastings algorithm is as follows:

1. Choose a random initial starting point in parameter space, $\boldsymbol{\theta}_0$, where $\boldsymbol{\theta}$ is the vector of parameters being estimated.

   Repeat N times:

2. Take a random jump in parameter space such that $\boldsymbol{u} = \boldsymbol{\theta}_i + \Delta\boldsymbol{\theta}$ where $\Delta\theta \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\sigma})$ and i is the step number. The random jump can be drawn from any proposal distribution, but a properly tuned proposal distribution results in much faster convergence to the posterior and a multivariate normal distribution is easy to sample from and is effective for many classes of problems[81]. The vector, $\boldsymbol{\sigma}$, is the vector of standard deviations for each parameter and is chosen before starting the chain.

3. Calculate the value
$$R = \frac{\mathcal{L}(\boldsymbol{u})}{\mathcal{L}(\boldsymbol{\theta})}\frac{P(\boldsymbol{u})}{P(\boldsymbol{\theta})}, \tag{3.12}$$
   where $\mathcal{L}$ is the likelihood and P is the prior.

4. Accept the proposed step $\boldsymbol{u}$ with probability $min(1, R)$.

5. If the step is accepted, set $\boldsymbol{\theta}_{i+1} = \boldsymbol{u}$, otherwise set $\boldsymbol{\theta}_{i+1} = \boldsymbol{\theta}_i$.

6. Record the value of $\boldsymbol{\theta}_{i+1}$ to the chain.

   End repeat.

7. To obtain the marginalised posterior for any parameter(s), simply histogram the column(s) of the chain corresponding to that parameter(s).

Burn-in, thinning and convergence This algorithm is guaranteed to converge to the true posterior given an infinite amount of time, however, in practice, the number of steps, N, is a

finite number and the chain must be checked for convergence. Since the chain generally starts far from the peak of the posterior, the initial part of the chain, called the "burn-in" is not an accurate reflection of the posterior and is generally removed [82]. Since the points in the chain are correlated, it can appear to have converged when it hasn't. This can be alleviated by "thinning", which is the removal of (for example) every third or fourth point of the chain. After these have been performed, the chain should be checked for convergence. Several tests have been derived to do this, including tests based on spectral analysis or correlations between chains [83], but one of the most common is the Gelman-Rubin [84] test.

The Gelman-Rubin criterion for convergence is defined as follows: let M be the number of chains, n the number of steps in a chain after burn-in is removed (assumed to be the same for all chains), $\bar{\theta}_m$ the mean of the parameter estimates for chain m and $\bar{\theta}$ the mean of the parameters for all the chains combined (here, θ is understood to be a vector and $\bar{\theta}$ is a vector of means). Then let

$$B = \frac{n}{M-1}\sum_{m=1}^{M}(\bar{\theta}_m - \bar{\theta})^2 \tag{3.13}$$

and

$$W = \frac{1}{M}\sum_{m=1}^{M} s_m^2 \quad \text{where} \quad s_m^2 = \frac{1}{1-n}\sum_{t=1}^{n}(\theta_m^t - \bar{\theta}_m)^2. \tag{3.14}$$

B is the between-chain variance and W is within-chain variance. The posterior marginal variance, $\hat{V}$, is the weighted average of W and B:

$$\hat{V} = \frac{n-1}{n}W + \frac{M+1}{nM}B. \tag{3.15}$$

If all the chains have converged, the posterior marginal variance should be close to the within-chain variance, i.e. $\hat{V}/W \approx 1$.

Proposal distribution

The choice of the proposal distribution affects the rate of convergence of the chain. If the proposal distribution is too "narrow" in parameter space the chain will take a long time to explore the entire posterior. If it is too large, the chain will also struggle to map out the posterior. In Metropolis-Hastings, the proposal distribution must be fixed before the chain starts running to ensure convergence to the true posterior, thus it can take several attempts to find an acceptable proposal distribution. The acceptance ratio, defined as the ratio between the number of successful steps and the total number of steps, can be used as a diagnostic to

determine whether the proposal distribution has been well chosen. The acceptance ratio should be anywhere between 20% and 60%.

Diagonalising the covariance matrix

Usually, parameters are correlated, meaning that the Gaussian proposal distribution described in the above algorithm will be inefficient. The correlations between parameters can be described by the covariance matrix: if an MCMC chain, X, is represented as a matrix with N rows and n columns (N being the number of steps and n being the number of parameters), then the elements of the covariance matrix, C, are given by:

$$C_{ij} = \sum_{k=1}^{N} \frac{(X_{ki} - \langle X_i \rangle)(X_{kj} - \langle X_j \rangle)}{N}, \tag{3.16}$$

where $\langle X_i \rangle$ indicates the mean of the i'th column of the chain. This covariance matrix can be used to make the algorithm take correlated jumps in parameter space. Then the new jump in step 2 of the above algorithm becomes

$$\Delta\boldsymbol{\theta} = Q \cdot \boldsymbol{\zeta} \tag{3.17}$$

where Q is the Cholesky decomposition [85] of the covariance matrix and $\boldsymbol{\zeta}$ is a vector of random numbers drawn from a Gaussian of zero mean and unit variance.

3.4 Bayesian Model Comparison

Often in science, we are not merely interested in the best fitting parameters of a model, but also how that model compares with other models. There are two competing factors at play in model comparison: goodness of fit and model complexity. A model with more degrees of freedom may be a better fit to the data, but those extra parameters may be undesirable. The guiding principle for scientific theories can be summarised by Einstein's quote: "Everything should be as simple as it can be, but not simpler". This can be seen as a rephrasing of Occam's razor[1] - the most preferred model is the simplest one which is compatible with the evidence. Bayesian model comparison offers a framework in which extra model complexity is penalised, if it is not required by the data. We will define a model, $\mathcal{M}$, as a set of parameters, θ, and their prior distribution, $P(\theta|\mathcal{M})$. It is important that the prior depends only on the model constraints,

[1] William of Occam (ca. 1285-1349): "Pluralitas non est ponenda sine neccesitate".

not on previous experiments, in order to evaluate the economy of the model since a large range of parameter values implies increased model complexity.

3.4.1 The Bayesian Evidence

The Bayesian evidence, also known as the marginal likelihood or the model likelihood, is essential for evaluating a model's performance given the data [76]. It is given by:

$$P(D|\mathcal{M}) = \int P(D|\theta, \mathcal{M})P(\theta|\mathcal{M})d\theta, \tag{3.18}$$

where it is understood the integral is over the full parameter space of θ. It should be noted that this equation is simply the normalisation condition of Eq.(3.7): as the posterior distribution is a probability, it must integrate to one over all parameter space. The Bayesian evidence can be seen as the average of the likelihood, under the prior for a specific model choice [76]. Using Bayes' Theorem, we can derive the model posterior probability:

$$P(\mathcal{M}|D) \propto P(\mathcal{M})P(D|\mathcal{M}), \tag{3.19}$$

where we have ignored the constant of proportionality, $P(D)$, since it depends only on the data. $P(\mathcal{M})$ is the prior probability assigned to the model and is usually taken to be non-committal, i.e. if there are N_m models, we take $P(\mathcal{M}) = 1/N_m$ for each model.

We can use Eq.(3.19) to directly compare two models, $\mathcal{M}_1$ and $\mathcal{M}_2$, by considering the ratio of posterior probabilities, or "posterior odds":

$$\frac{P(\mathcal{M}_1|D)}{P(\mathcal{M}_2|D)} = B_{12}\frac{P(\mathcal{M}_1)}{P(\mathcal{M}_2)}, \tag{3.20}$$

where B_{12} is the Bayes factor and is given by

$$B_{12} = \frac{P(D|\mathcal{M}_1)}{P(D|\mathcal{M}_2)}. \tag{3.21}$$

If $B_{12} > 1$, the data favour model 1 over model 2, and vice versa. The Bayes factor represents by how much the relative odds of the two models have changed in light of the new data, regardless of our previous beliefs in the models.

Bayes factors can be interpreted using the empirically calibrated Jeffreys' scale, which is displayed in Table 3.1. Given a value for the Bayes factor, the Jeffreys' scale indicates how strongly the evidence favours one model over another. It is often represented in terms of odds (the third column of the table), which are easier to understand intuitively than the Bayes factors themselves (most will agree that $150:1$ are strong odds, whereas $3:1$ is less conclusive).

TABLE 3.1: Jeffreys' scale: the empirically determined scale to evaluate how strongly model 1 is favoured over model 2. Adapted from [76].

$\lvert\ln B_{12}\rvert$	Odds	Probability	Strength of Evidence
< 1.0	$\lesssim 3:1$	< 0.750	Inconclusive
1.0	$\sim 3:1$	0.750	Weak Evidence
2.5	$\sim 12:1$	0.923	Moderate Evidence
5.0	$\sim 150:1$	0.993	Strong Evidence

3.5 Fisher Matrix Analysis

When designing an experiment, we may wish to know how the design of the experiment affects the expected errors on the parameters constrained by the experiment. While one could generate many realisations of mock data and analyse them with MCMC or a similar technique, the Fisher matrix [86] allows one to compute the errors without any data required, under the assumption that the likelihood is a multivariate Gaussian. A Fisher matrix analysis is essentially a way of doing propagation of errors with correlated data and many parameters [87]. The Fisher matrix allows us to estimate not only the error on each parameter, but also the correlations between them.

Let $\boldsymbol{\theta}$ be a vector of parameters and $\boldsymbol{\theta}^*$ be the fiducial model. We can expand the likelihood about $\boldsymbol{\theta}^*$ [87]:

$$\begin{aligned}\ln\mathcal{L}(\boldsymbol{\theta}^* + \delta\boldsymbol{\theta}) = \ln\mathcal{L}(\boldsymbol{\theta}^*) \quad &+ \sum_i \left.\frac{\partial \ln\mathcal{L}(\boldsymbol{\theta})}{\partial\theta_i}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \delta\theta_i \\ &+ \frac{1}{2}\sum_{ij} \left.\frac{\partial^2 \ln\mathcal{L}(\boldsymbol{\theta})}{\partial\theta_i\partial\theta_j}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \delta\theta_i\delta\theta_j \\ &+ \frac{1}{6}\sum_{ijk} \left.\frac{\partial^3 \ln\mathcal{L}(\boldsymbol{\theta})}{\partial\theta_i\partial\theta_j\partial\theta_k}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \delta\theta_i\delta\theta_j\delta\theta_k + \cdots . \end{aligned} \tag{3.22}$$

The first term is a constant which depends only on the fiducial model. Since, after many data realisations, we expect the fiducial model to be the point of maximum likelihood, then by

definition the second term will vanish. The third term is the curvature matrix or Hessian of the likelihood and defines the Fisher matrix [86]:

$$F_{ij} = \left\langle -\frac{\partial^2(\ln\mathcal{L})}{\partial\theta_i\partial\theta_j} \right\rangle . \tag{3.23}$$

The angle brackets represent the expectation value which, for an arbitrary function $g(x)$, is defined as:

$$\langle g(x) \rangle = \int_{-\infty}^{\infty} g(x) f_X(x) dx, \tag{3.24}$$

where $f_X(x)$ is the probability distribution function for the random variable x, which is usually the noise on the data, assumed to be Gaussian with mean zero.

To write the Fisher matrix in terms of the theoretical predictions for the observables, $\boldsymbol{X}$, we first note that the likelihood is given by:

$$\mathcal{L} = \frac{1}{\sqrt{(2\pi)^N |\boldsymbol{C}|}} \exp\left(-\tfrac{1}{2}\boldsymbol{\Delta}^T \boldsymbol{C}^{-1} \boldsymbol{\Delta}\right), \tag{3.25}$$

where $\boldsymbol{\Delta} = \boldsymbol{X} - \boldsymbol{d}$ (the difference between the theoretical prediction and the measured quantity), $\boldsymbol{C}$ is the data covariance matrix and N is the number of data points. Substituting this into Eq.(3.23), we find:

$$F_{ij} = \frac{\partial \boldsymbol{X}^T}{\partial\theta_i} \boldsymbol{C}^{-1} \frac{\partial \boldsymbol{X}}{\partial\theta_j} + \frac{1}{2}\left(\boldsymbol{C}^{-1}\frac{\partial \boldsymbol{C}}{\partial\theta_i}\boldsymbol{C}^{-1}\frac{\partial \boldsymbol{C}}{\partial\theta_j}\right). \tag{3.26}$$

If the data covariance matrix, which also includes effects like cosmic variance, is independent of the parameters, as is often the case [87], and if the data are uncorrelated, the Fisher matrix becomes:

$$\begin{aligned} F_{ij} &= \frac{\partial \boldsymbol{X}^T}{\partial\theta_i} \boldsymbol{C}^{-1} \frac{\partial \boldsymbol{X}^T}{\partial\theta_j} \\ &= \sum_n \frac{1}{\sigma_n^2} \frac{\partial X_n}{\partial\theta_i} \frac{\partial X_n}{\partial\theta_j}, \end{aligned} \tag{3.27}$$

where n is the index over the data.

For multiple, independent datasets, the Fisher matrix for each dataset can simply be added together to produce a total Fisher matrix. Additionally, any prior information (for example, constraints from previous surveys) can be encoded into a prior matrix, which can also be added to the Fisher matrix to affect the total information content.

Whereas marginalisation over nuisance parameters requires an MCMC calculation in general, it can be done trivially with the Fisher matrix, under the assumption that the likelihood is a multivariate Gaussian. The Fisher matrix can be written in terms of sub-matrices [87]:

$$F = \begin{pmatrix} \Phi & \mathcal{O} \\ \mathcal{O} & \Psi \end{pmatrix}, \tag{3.28}$$

where Φ is the sub-matrix corresponding to the parameters of interest, the matrix $\mathcal{O}$ describes the correlations between the nuisance parameters and the parameters of interest and Ψ is the sub-matrix corresponding to the nuisance parameters. The marginalised Fisher matrix is given by [88]:

$$\widetilde{F} = \Phi - \mathcal{O}\Psi^{-1}\mathcal{O}^T, \tag{3.29}$$

where the first term represents the matrix of the reduced parameter space of interest and the second term corresponds to the effect of marginalisation over the nuisance parameters.

The inverse of the Fisher matrix estimates the parameter covariance matrix. In the case for an unbiased estimator (i.e. the expected value of $\boldsymbol{\theta}$ corresponds to $\boldsymbol{\theta}^*$) and where one does not marginalise over any parameters, then the expected error satisfies the Cramér-Rao bound [89]:

$$\Delta\theta_i \geq \frac{1}{\sqrt{F_{ii}}}. \tag{3.30}$$

In the more realistic case which includes marginalisation, this inequality becomes:

$$\Delta\theta_i \geq \sqrt{F^{-1}_{ii}}. \tag{3.31}$$

It can be shown that the error obtained when marginalising is always greater than or equal to the error obtained without marginalisation. It should also be noted that in the case where the likelihood is exactly Gaussian, the Cramér-Rao bound becomes an equality.

As the likelihood is assumed to be a multivariate Gaussian for a Fisher matrix analysis, the resulting error contours are ellipsoids (ellipses for two parameters), given by the equation

$$\Delta\boldsymbol{\theta}^T \widetilde{F} \Delta\boldsymbol{\theta} = \beta, \tag{3.32}$$

where $\Delta\boldsymbol{\theta} = \boldsymbol{\theta} - \boldsymbol{\theta}^*$ and β is a constant determined by the desired confidence level and the number of parameters. In the two parameter case, $\beta = 2.31$ and 6.17 for the 1 and $2-\sigma$ contours respectively [88]. An example of the effect of the the fiducial model on error ellipses is shown in Figure 3.2 for the CPL parameterisation of dark energy [90, 91]. As w_0 and w_a vary, not only does the centre of the ellipse change but the size and slope of the degeneracy direction changes as well.

In order to evaluate the amount of information obtained for a given Fisher matrix analysis and directly compare surveys, various figures of merit have been developed, such as the volume of the error ellipsoid, the $1-\sigma$ area of the two dimensional ellipse (for two parameters), the trace of the inverse marginalised Fisher matrix and the sum of the elements of the inverse marginalised Fisher matrix [87]. These figures of merit are not necessarily equivalent and can affect the choice of survey design based on a Fisher matrix analysis.

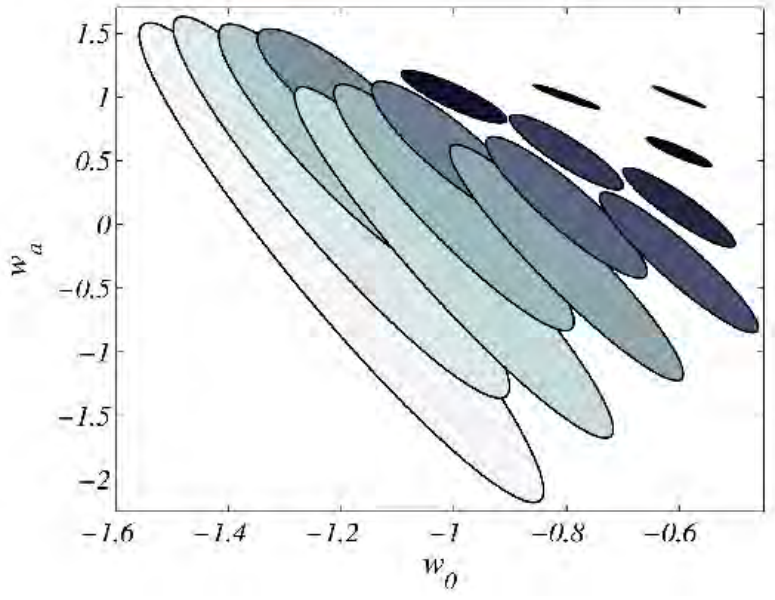


FIGURE 3.2: Illustration of the effect of changing the fiducial values for w_0 and w_a (from the CPL parameterisation) on the error ellipses for a mock survey supplying Hubble parameter, $H(z)$ and angular diameter distance $d_A(z)$ data. w_0 and w_a have been varied on a grid over $-1.3 < w_0 < -0.6$, $-0.7 < w_a < 1$. The centre, size and the slope of the degeneracy direction changes with w_0 and w_a. It can be seen that the area of the ellipse decreases for increasing w_0 and w_a implying that these parameters can be measured more precisely at higher values. Taken from [87].

Chapter 4

Extending the BEAMS Formalism

4.1 Introduction

As discussed in Section 1.7.1, Type Ia supernovae (SN Ia) are standardisable candles, making them one of the most reliable distance measures and a cornerstone of cosmology ever since the discovery of the late time accelerated expansion of the Universe [26, 27].

Future surveys which will produce large amounts of photometric data, such as the Dark Energy Survey (DES) [92], Pan-STARRS [93] and the Large Synoptic Survey Telescope (LSST) [94], will increase the number of SN Ia candidates by orders of magnitude. While a foolproof method of identifying a Type Ia is to analyse its observed spectrum, taking spectra is expensive and, for surveys such as those mentioned above, it will be unfeasible to perform spectroscopic follow-up for all candidates, introducing a possible bias due to contamination from Type Ib/c and Type II supernovae, which we collectively denote non-Ia supernovae (SN nIa) [46, 95]. However, using the photometric information gathered by the survey, one can fit template light curve models to the data using a template fitter such as MLCS2k2 [96] or SALT2 [97] which gives a probability for each object to be a Type Ia [98].

As an example, Figure 4.1 shows supernova colours for different types of supernovae, illustrating that there are differences between the types and suggesting that the probability of an object being a Ia can be determined from the photometric data alone [99]. Figure 4.2 shows a sample of SDSS supernova data with type probability information.

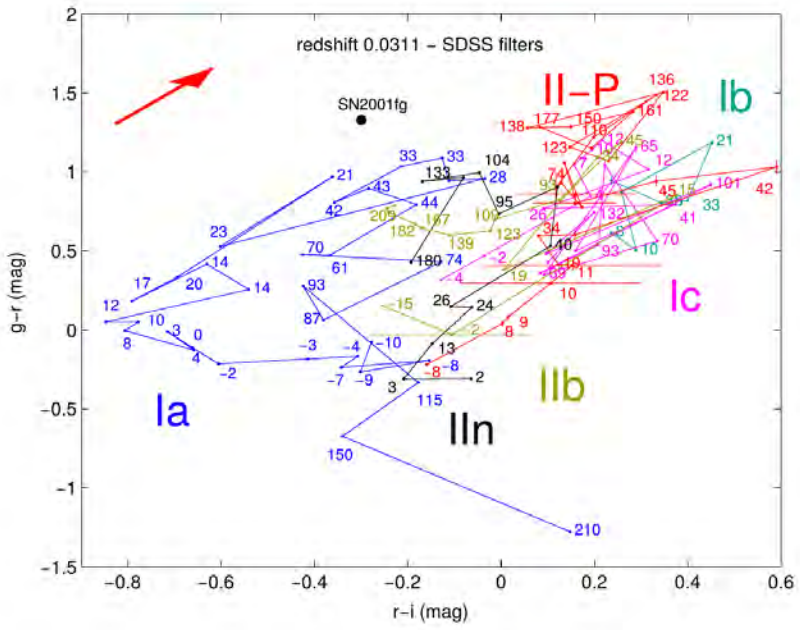


FIGURE 4.1: *Obtaining type probabilities from supernova colors:* Supernova colours (the magnitude in one colour filter subtracted from another) are shown for different supernova types at $z = 0.0311$. The time evolution in days for each type is depicted by a linearly connected track between points. In this particular case, SN2001fg was compared and found it was most likely a SN Ia, about 1 month past maximum brightness, as confirmed by spectroscopic follow-up [99].

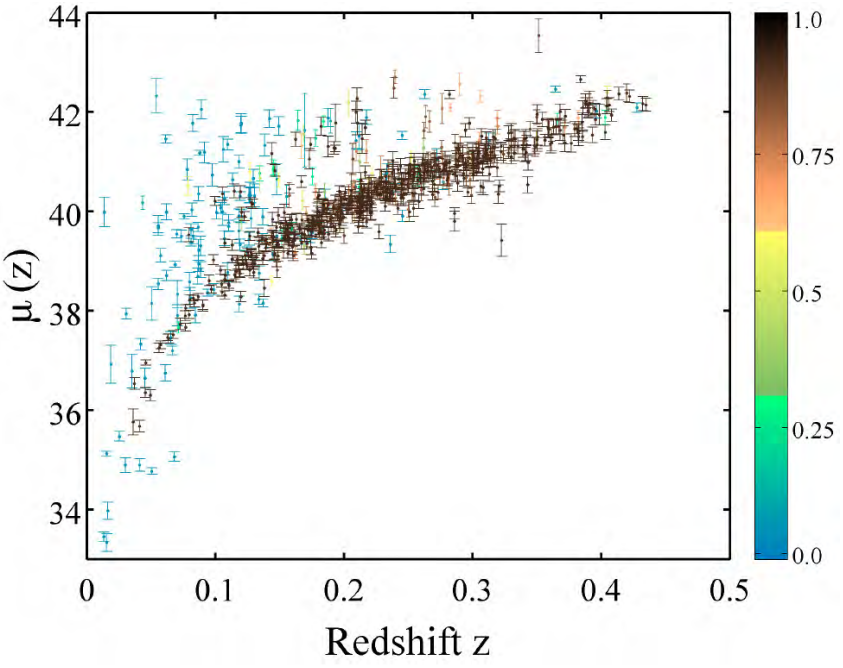


FIGURE 4.2: A sample of 792 SDSS-II supernovae, colour coded with probabilities from the PSNID typer [98, 100], used in the BEAMS analysis [46].

Previously, these SN Ia probabilities were used only to determine candidates for spectroscopic follow-up [100]. Without spectroscopic follow-up, applying a probability cut (for example, taking all supernovae with probability greater than 0.9 to be a Type Ia) will introduce a bias in the cosmological parameters [46]. To avoid such biases, one can either demand a very high purity, which excludes much of the data [101, 102] or use all the data within a statistical framework that accounts for the contamination. One such method, developed by Kunz, Bassett & Hlozek [45] is the Bayesian Estimation Applied to Multiple Species (BEAMS). BEAMS has recently been applied to the full three years of data from the SDSS-II supernova survey [100, 103], which reduced the $\Omega_m - \Omega_\Lambda$ contours by a factor of three relative to the spectroscopic data alone [46].

Despite this success, the current implementation of BEAMS assumes the supernova errors are not correlated with each other; an approach which will not be appropriate for future surveys. To analytically account for correlations between supernovae errors in the BEAMS posterior requires summing over m^N terms, where N is the number of supernova candidates and m is the number of possible supernova types, which here we take to be 2, corresponding to Ia's and nIa's. Clearly, this is computationally impossible, but in this paper we will show that if the form of the covariance matrix is known, BEAMS can still be used to estimate cosmological parameters in an unbiased way, using a numerical marginalisation over supernova type (see Section 4.4) [1].

4.2 BEAMS

Cosmological parameter estimation usually proceeds by maximising the posterior, $P(\theta|D)$ (see Chapter 3), where D is the set of redshifts and distance moduli of spectroscopically confirmed Type Ia supernovae and θ is the set of cosmological parameters, such as Ω_m, Ω_Λ and H_0. What happens if we do not have spectroscopic confirmation of an object's type but only a probability that it is a Ia? Unbiased parameter estimation in this case can be achieved using Bayesian Estimation Applied to Multiple Species (BEAMS).

BEAMS [45] considers all data in a given sample, appropriately weighting the likelihood of each data point based on its probability of being a Ia. Let τ_i be the type of object i and $\tau_i = \text{Ia}$ if the object is a Type Ia and $\tau_i = \text{nIa}$ if the object is not a Ia (for example, if it is a Type Ib/c or a Type II supernova). Then the posterior, $P(\theta|D)$ (here θ, D and τ are understood to be vectors), is:

$$P(\theta|D) = \sum_{\tau} P(\theta, \tau|D). \tag{4.1}$$

This sum marginalises over all possible combinations of types for the dataset so τ here is a length-N vector (where N is the number of objects). For example, if there were three objects in the dataset, the first term in the sum would have $\tau_1 = \text{Ia}$, $\tau_2 = \text{Ia}$, $\tau_3 = \text{Ia}$, the second term would have $\tau_1 = \text{nIa}$, $\tau_2 = \text{Ia}$, $\tau_3 = \text{Ia}$ etc.

Thus, for the case of two distinct object types, this is a 2^N summation. Applying Bayes' theorem (Eq.(3.7)) gives that

$$P(\theta, \tau|D) = P(D|\theta, \tau)\frac{P(\theta, \tau)}{P(D)}. \tag{4.2}$$

$P(D)$, the Bayesian evidence, can be considered as a normalisation factor and ignored in further calculations. We will assume that $P(\theta, \tau) = P(\theta)P(\tau)$, since in the case of supernovae, $\mathrm{P}(\theta)$ depends on large scale cosmological evolution while $\mathrm{P}(\tau)$ depends on local gastrophysics [45]. $P(\theta)$ is the usual prior on the parameters (probability based on prior knowledge about the parameters), and $P(\tau)$ can be written as

$$P(\tau) = \prod_{\tau_j=\text{Ia}} P_j \prod_{\tau_k=\text{nIa}} (1 - P_k). \tag{4.3}$$

This is the product of the probabilities, P_i, for all the objects typed as a Ia, $\tau_i = \text{Ia}$, multiplied by the product of $(1 - P_i)$ for all the objects typed as nIa's, $\tau_i = \text{nIa}$. This assumes the data and the object types are uncorrelated (see [47] for details of BEAMS without this assumption, which is also briefy discussed below).

Thus the BEAMS posterior is given by

$$P(\theta|D) \propto P(\theta) \sum_{\tau} \left[P(D|\theta, \tau) \prod_{\tau_j=\text{Ia}} P_j \prod_{\tau_k=\text{nIa}} (1 - P_k) \right]. \tag{4.4}$$

As an order 2^N calculation, this is computationally unfeasible. In the case of uncorrelated data, there is a simplification we can use to make the problem tractable. For any independent data, the joint likelihood can always be written as a product of likelihoods. Thus,

$$P(D|\theta, \tau) = \prod_{\tau_i=\text{Ia}} P(D_i|\theta, \tau_i = \text{Ia}) \prod_{\tau_\text{j}=\text{nIa}} \mathrm{P}(\mathrm{D_j}|\theta, \tau_\text{j} = \text{nIa}). \tag{4.5}$$

So the posterior is now a sum over products. This can be further simplified by applying the fact that any binomial combination can be written as a product of sums of two terms:

$$\sum_{\tau} \prod_{\tau_i=\mathrm{Ia}} A_i \prod_{\tau_j=\mathrm{nIa}} B_j = \prod_k (A_k + B_k). \tag{4.6}$$

In the case of supernova data, A_k is simply the likelihood that object k is a Type Ia, $\mathcal{L}_{\mathrm{Ia,k}}$, multiplied by P_k, and similarly, B_k is the likelihood that object k is a nIa, $\mathcal{L}_{\mathrm{nIa,k}}$, multiplied by $(1 - P_k)$.

Finally, the uncorrelated form of the BEAMS posterior is written as

$$P(\theta|D) \propto P(\theta) \prod_{i=1}^{N} \Big(\mathcal{L}_{\mathrm{Ia,i}}(\theta) P_i + \mathcal{L}_{\mathrm{nIa,i}}(1 - P_i) \Big). \tag{4.7}$$

The form of the likelihood for a SN Ia, $\mathcal{L}_{\mathrm{Ia,i}}$, is well defined as

$$\mathcal{L}_{\mathrm{Ia,i}} = \exp \frac{-(\mu_i - \mu_{\mathrm{Ia}}(\theta, z_i))^2}{2\sigma_i^2}, \tag{4.8}$$

where μ_i is the distance modulus estimate (see Eq.(1.38)) for the object, $\mu_{\mathrm{Ia}}(\theta, z_i)$ is the theoretical prediction and σ_i is the uncertainty for that data point. However, nIa's do not have the same intrinsic magnitude distribution as Ia's and moreover, their distribution is very broad meaning little, if any, cosmological information can be determined from them. So what form should the nIa likelihood take? In Section 4.2.1 we show that as long as the form of the nIa likelihood is reasonably general, it is possible to fit for its parameters in the analysis and produce unbiased results.

Eq.(4.7) works well for uncorrelated data but is unlikely to be accurate in the case of correlated data. For the large datasets which will be available in the future, correlations among the data cannot be ignored and another solution must be found to apply BEAMS to these datasets in a computationally feasible way. Section 4.4 demonstrates possible solutions to the 2^N 'exponential catastrophe'.

Prompted by the assumption made in Kunz *et al.*[45] that the data and probabilities are independent, Newling *et al.*[47] discussed an extension to the BEAMS formalism, in addition to studying the importance of factors such as sample size and biases in the probabilities. They rewrite the BEAMS posterior in a form where the data are conditional on the probabilities of the objects. In the case of supernovae, the distribution of distance moduli of the objects may

be dependent on the probabilities if the types are unknown, because we expect SN Ia's to be brighter than nIa's (so we expect an object with a high probability to have a smaller distance modulus). The strength of the correlation between distance modulus estimates and probability depends on how the probabilities are derived (light curve fitting, using the Hubble diagram directly etc.).

However, it has been shown [47] that a more general form of the posterior, which includes this possible correlation, can be used. This extended form of the posterior is narrower in comparison to the the posterior of Eq.(4.4), in the presence of correlations between the data and the probabilities. Newling *et al.* also showed that if there are biases in the probabilities, it is possible to still obtain accurate estimates of the cosmological parameters using a debiasing technique, if all the variables used when estimating the probabilities are known.

4.2.1 Testing BEAMS

Kunz *et al.*[45] performed several tests of the BEAMS method, comparing it with other methods. Figure 4.3 shows the bias and error (on the estimated cosmological parameters) for several methods using mock supernova data. Only BEAMS is able to produce an unbiased estimate of the parameters with a small error. Kunz *et al.* also considered the case where the probabilities are unknown, are uncertain or are globally biased. They showed that a global bias in the probabilities can be corrected by including a shift parameter in the likelihood and showed that uncertainties in the probabilities which are expected from light curve fitters can be dealt with in the BEAMS formalism.

4.3 Applying BEAMS

Hlozek *et al.*[46] performed extensive tests on simulated data to demonstrate the effectiveness of BEAMS. They constructed a dataset of simulated distance moduli and redshifts for Ia's and nIa's drawn from Gaussian distributions, as well as a dataset using the supernova simulation package, SNANA [104]. These simulations were both analysed assuming a ΛCDM cosmology and fitting for the parameters Ω_m, Ω_Λ and H_0. As the nIa likelihood is unknown, they assumed a Gaussian distribution for the nIa's with a mean that is allowed to evolve with redshift. All parameters in the nIa likelihood are fit for in the analysis.

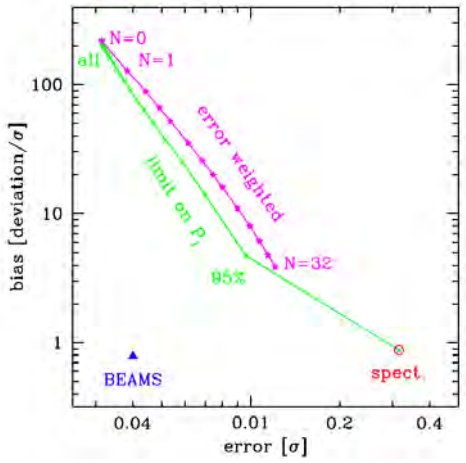


FIGURE 4.3: Comparison of BEAMS with using only the spectroscopic data (red circle), a probability cut (green line) and a weighted χ^2 method where points with higher probability have greater error (pink line, increasing N means a larger penalisation for low probabilities). Only BEAMS and the spectroscopic subsample are unbiased. BEAMS reduces the error by a factor of 8 owing to the utilisation of the full dataset [45].

The resulting contours for both simulations can be seen in Figure 4.4. There is a small ($\sim 1\sigma$) bias in the BEAMS contours from the SNANA results which could be due to a number of factors including: modeling the nIa's as only one population, incorrect treatment of Malmquist bias (we expect each population to require a different correction due to Malmquist bias) and possible errors in the SNANA simulations. A bias between spectroscopic and BEAMS results is not observed when using real data.

Figure 4.5 shows the results when BEAMS is applied to the SDSS dataset of photometric SNe. This dataset consists of 792 supernovae, with only 297 objects spectroscopically confirmed as Type Ia. The full dataset represent a sample of high quality photometric data, which have host galaxy redshifts[1]. Unlike in the test cases, when a probability cut is applied to the SDSS data, Figure 4.5 does not show a bias. This is most probably because the SDSS dataset is much smaller than the mock cases considered and the biases will only become obvious in future datasets of thousands of supernovae. The important thing to note in the results is that the BEAMS contours are a factor of three smaller than the spectroscopic data contours, and are still unbiased. This is because BEAMS can use the Ia's that one would normally have to throw away. This shows that BEAMS can produce tighter cosmological constraints from supernova measurements, while remaining unbiased.

[1]It is of course far easier to obtain a host galaxy redshift than to do spectroscopic follow-up of a supernova because supernovae are temporary while their hosts persist and massive galaxy spectroscopic surveys are becoming more and more common. Still, we believe it is possible to do BEAMS with only photometric host galaxy redshifts by marginalising over the uncertainty in redshift for each supernova. This will likely be a project for the BEAMS group.

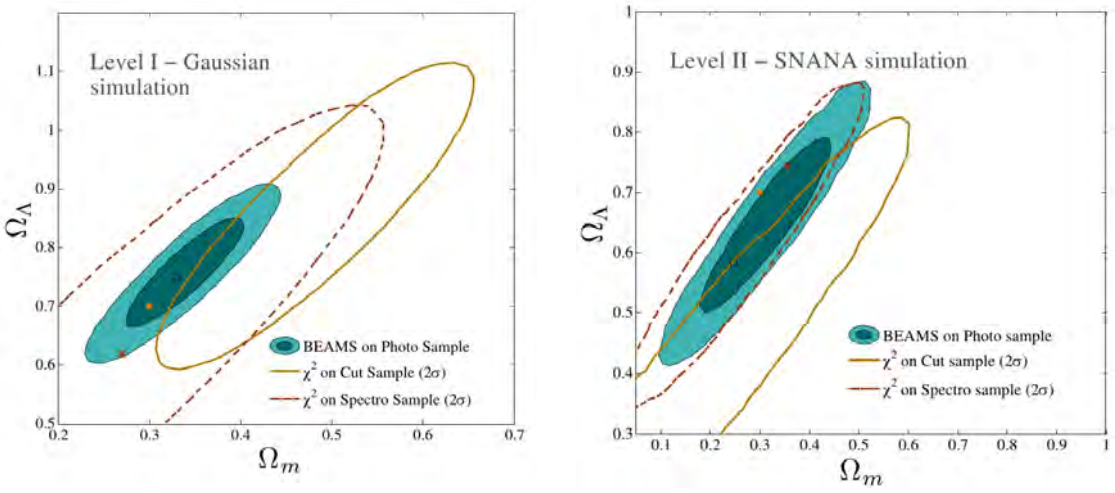


FIGURE 4.4: $\Omega_m - \Omega_\Lambda$ contours obtained when analysing simulated Gaussian data (left panel) and SNANA simulated data (right panel). The dashed line contours are obtained from the subsample of spectroscopically confirmed supernovae, the lined contours from a subsample where $P_i > 0.9$ for each object and the filled contours from all the data analysed using BEAMS. The input cosmology is given by an orange square, the best fit BEAMS cosmology by an empty square and the best fit cosmology from the spectroscopic sample by a red cross. It is clear that the spectroscopic subsample gives unbiased results at the cost of reducing the sample size while the probability cut results in biased contours. BEAMS is able to recover the true cosmology with substantially smaller error than the spectroscopic sample.

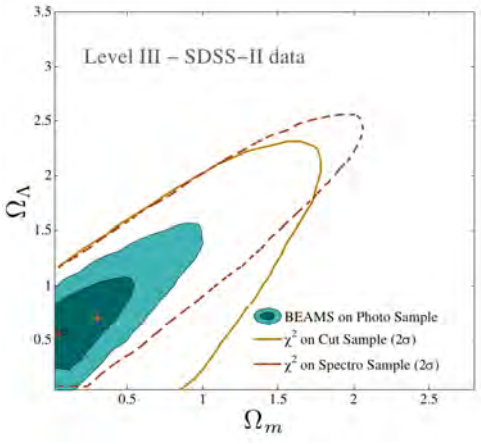


FIGURE 4.5: As in Figure 4.4, the dashed line contours are obtained from the subsample of spectroscopically confirmed supernovae, the lined contours from a subsample where $P_i > 0.9$ for each object and the filled contours from all the data analysed using BEAMS. Although of course the true cosmology is not known, the current fiducial cosmology is given by an orange square, the best fit BEAMS cosmology by an empty square and the best fit cosmology from the spectroscopic sample by a red cross. In this case, the cut sample is unbiased relative to the spectroscopic sample, but this is most likely because this dataset is much smaller than the mock datasets (there are only 792 total SNe). However, one should note that the BEAMS contours are a factor of three smaller than the spectroscopic data contours.

4.4 Extending BEAMS to Correlated Data

4.4.1 Correlated Supernova Data

Correlations between supernova data has only recently become an issue that must be included in cosmological parameter sets (e.g. [22, 42]) but will become progressively more important as we push on the systematics floor related to SN Ia surveys. Correlated systematic uncertainties, the focus of this work, can arise from a large number of sources and a detailed study of correlations has yet to be undertaken due to the diversity and complexity of the various contributing factors. Schematically correlations can arise from:

- *Peculiar velocities:* When supernovae are within 50 Mpc or so of each other the peculiar velocities of their host galaxies will be correlated by large-scale bulk flows. These peculiar velocities cause correlated redshift errors. Usually redshift errors are converted into additional distance modulus errors (see e.g. [105]) but even if this is not done it will cause errors in the recovered cosmological parameters. See, for example, [106–109].

- *Redshift-colour and redshift-stretch correlations:* If there is no spectroscopic host galaxy redshift for an object, the redshift is estimated photometrically either from the host or the supernova multi-band light curves. The effect of redshift on the light curve is degenerate to some extent with the stretch and colour corrections. Hence errors on redshift will correlate with those in the colour and stretch and thus with the estimated distance modulus of the supernova [110].

- *Filter errors:* Transmission curves for the actual filters used on a telescope approximate true through-put. Errors in measuring these transmission curves (or time dependent changes of the filters) [111] will tend to induce redshift-dependent correlations. For example, in [112], the authors consider the error in the flux calibration of the telescope which, since it affects each filter differently, correlates objects at similar redshifts. Zero-point photometry errors are a major source of uncertainty in current surveys that are common to all supernovae observed with the same telescope; e.g. [42].

- *Template error correlations:* Unaccounted for evolution of supernovae with redshift causes correlated errors due to errors in the light curves that form the training set. An example of this is the 'U-band anomaly' which causes discrepancies between the SALT2 and MLCS2k2

light curve fitters which may be related to an excess of flux in the UV at high redshift in SN Ia [105, 113].

- *Observational conditions:* Bad weather will cause holes in the light curve coverage of all supernovae visible at a given time, while seeing conditions will alter photometry measurements in a correlated way. These will induce subtle correlations between measurements of objects observed on the same night in similar conditions [94, 114].

- *Combining data from multiple telescopes:* The covariance matrix for combining data from multiple telescopes can be very complex, as discussed in the 3 year SNLS analysis [22]. With the exception of perhaps the final LSST dataset, combining data from multiple surveys will continue to be standard.

- *Gravitational lensing:* Supernovae that are close together on the sky, at similar redshifts will experience similar brightening or dimming due to lensing, depending on the matter distribution along the line of sight. Future large surveys will have mass maps and hence will be able to predict and remove this signal to some extent [115–117].

- *Dust:* Supernovae along neighbouring lines of sight will suffer similar extinction from the Milky Way and from any intergalactic dust, which will induce correlations [118, 119].

- *Host-galaxy correlations:* There is now solid evidence that dispersion in the Hubble diagram correlates with the properties of host galaxies, particularly the galaxy type, size and mass. See, for example [120–123].

- *Spectroscopic targeting correlations:* Since spectroscopic follow-up is typically not random, there may be hidden correlations. For example, follow-up may favour candidates well-separated from the host galaxy core. Malmquist bias can also cause correlations which depend on the details of the spectroscopic survey [100, 105]. If there is an unknown systematic that such objects are intrinsically brighter/fainter than average, this will cause a correlated systematic error [100].

- *nIa correlations:* Many of the sources of correlation listed here will also affect nIa's, causing correlations between their distance moduli. However, these correlations will typically be much smaller than the intrinsic dispersion of the nIa population.

Given the complexity of these various effects, developing a 'realistic' correlation matrix is beyond the scope of this work and will need to be laboriously built from simulations and detailed

studies. Here we wish instead to develop ways of dealing with the general problem posed by the 2^N 'exponential catastrophe' that correlations present. We use several toy model covariance matrices to study potential resolutions of this catastrophe, finding that numerical marginalisation over the supernova types performs best.

4.4.2 Mock Data

In order to determine the effect that correlated data can have on parameter estimation with BEAMS, we use mock supernova datasets, the properties of which are known. This enables us to estimate the magnitude of the bias introduced by correlations between the data points and to determine the optimum way to handle this additional source of error.

To simulate supernova data, we need to create a distance modulus $\mu(z)$ for each object, as in Eq.(1.38). Through the distance modulus, SN Ia can be used to constrain cosmological parameters. The luminosity distance depends on the cosmological parameters (assuming a ΛCDM model) by

$$d_L(z) = \frac{c(1+z)}{H_0\sqrt{-\Omega_k}}\sin\left(H_0\sqrt{-\Omega_k}\int\frac{dz'}{H(z')}\right), \tag{4.9}$$

where

$$H(z) = H_0\left(\Omega_m(1+z)^3 + \Omega_\Lambda + \Omega_k(1+z)^2\right)^{1/2}, \tag{4.10}$$

and the reader is reminded that H_0 is the Hubble constant, Ω_m is the energy density of matter, Ω_Λ is the energy density of dark energy and Ω_k is the energy density of curvature (all densities relative to the critical density) [3]. This single equation is equivalent to the three separate curvature cases often quoted in textbooks, since $\sinh(x) = -i\sin(ix)$ [20]. We used a ΛCDM model to generate the mock data with the following parameters: $H_0 = 70.4$ km/s/Mpc, $\Omega_m = 0.272$ and $\Omega_\Lambda = 0.628$ (values taken from [124]).

SN Ia typically have very little scatter in their distribution of distance moduli, while nIa's tend to be widely scattered. SN Ia also tend to be brighter than nIa's. As such, we model the mock data as drawn from two populations: the Ia distribution is a narrow Gaussian centred on the fiducial cosmological distance modulus, $\mu(z)$, with standard deviation $\sigma_{\rm Ia}$. The nIa distribution is a wide Gaussian, centred on $\mu(z) + b$, where b is a constant shift[2]. The nIa distribution has standard deviation $\sigma_{\rm nIa}$. While in general one could model each known type of supernova as a

[2]It was shown in [46] that BEAMS is able to reconstruct the correct redshift evolution for the nIa distribution. Here we assume, without loss of generality, that b is a constant shift.

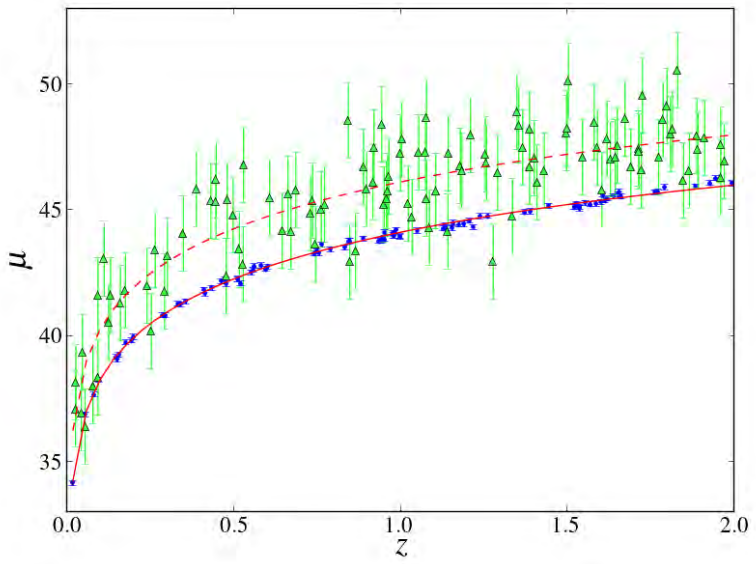


FIGURE 4.6: Distance modulus (μ) as a function of redshift for a typical mock dataset with 200 supernovae. Blue points are SN Ia, green triangles are SN nIa and the red line is the fiducial cosmological model. The dotted line is the model for the nIa's. The intrinsic dispersion of the Ia's is $\sigma_{\rm Ia} = 0.1$ and that of the nIa's is $\sigma_{\rm nIa} = 1.5$, as indicated by the error bars. We show the true types and intrinsic scatter of the points for clarity here but note that neither are given to BEAMS, which receives only the probabilities and the μ values as input.

separate population, two populations are sufficient for this work. For each object, redshifts in the range of $0.01 < z < 2.0$ were drawn from a uniform random distribution, while probabilities, P_i, were drawn from a distribution that simulates the expected Dark Energy Survey probability distribution, shown in Figure 4.7. There were roughly equal numbers of Ia's and nIa's. Here we assume that the probabilities are correct, but it was shown in [45, 46] that BEAMS can deal with biased probabilities in general.

The type of each object was randomly chosen with probability P_i which allowed the object to be assigned a distance modulus. Residuals were drawn from the appropriate distribution, depending on type, correlated using one of the covariance matrices described below and added to the appropriate Ia or nIa mean. A typical mock dataset distance modulus diagram is shown in Figure 4.6, where we have chosen the standard deviation of the Ia population to be $\sigma_{\rm Ia} = 0.1$, that of the nIa's to be $\sigma_{\rm nIa} = 1.5$ and the shift of the nIa population to be $b = 2.0$. In this paper we experimented with datasets varying between 200 and 1000 supernovae.

Since the exact form of the true covariance matrix for supernovae is unknown, we use simple toy models for a covariance matrix for the mock data. Our main goal is to show how BEAMS can be used for correlated data, so we do not attempt to model a realistic covariance matrix.

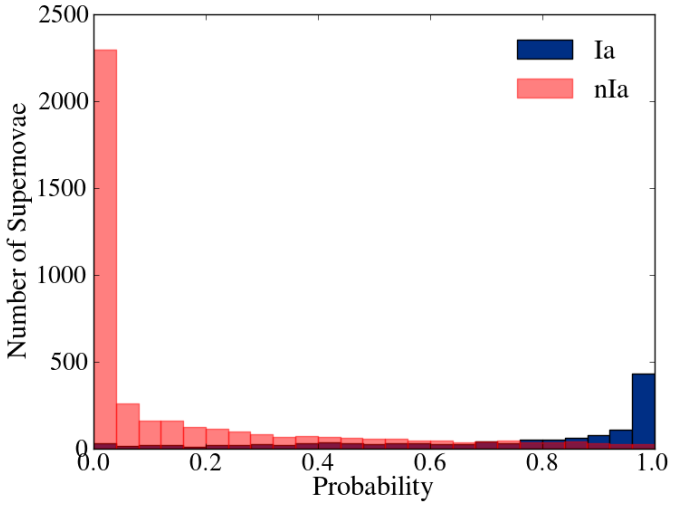


FIGURE 4.7: Histogram of Dark Energy Survey-like SN Ia probabilities, where the supernova data was generated using SNANA and fitted with MLCS2k2 to obtain the probabilities. The blue (dark), outlined bars are Ia's and the red (light) bars are nIa's. This figure illustrates the point that future data will deliver bimodal probabilities peaked near 0 and 1, and helps guide the probability distribution for our mock data.

We analyse the effects of three different covariance matrices. In all cases, we order the objects according to redshift, since we expect the strongest correlations to be functions of redshift.

As they will come from the same sources, we expect Ia-nIa and nIa-nIa correlations to be of similar magnitude as Ia-Ia correlations. However, since the dispersion of the nIa distribution is so large, these small correlations have little influence and only the Ia-Ia correlations affect the results. We found that when we created a dataset with correlations between all supernova types and analysed it assuming only Ia correlations, the results were identical to an analysis which included non-zero nIa-nIa and Ia-nIa correlations. Hence, we only consider Ia-Ia correlations in our analysis.

4.4.2.1 Wedding Cake Covariance Matrix

This example matrix is based on the supernova covariance matrix in [125], which is guaranteed to be positive-definite. This matrix has a wedding cake or step-like structure: an error entered at a given redshift contributes to the error of all higher redshift objects. This structure naturally arises as one loses observational features at higher redshifts thus introducing errors. We constructed a matrix based on this, in which we divided the data set into five redshift bins, adding an extra source of error in each bin. If each object can only be either of type Ia (τ_i=Ia)

or of type nIa (τ_i=nIa), the structure of this covariance matrix is given as:

$$C_{ij} = \sigma_i \sigma_j \delta_{ij} + V_{ij}, \tag{4.11}$$

where $\sigma_i = \sigma_{\rm Ia}$ if τ_i=Ia and $\sigma_i = \sigma_{\rm nIa}$ if τ_i=nIa and

$$V_{ij} = \begin{cases} \sum_{k=1}^{n_{i,j}} s_k & \text{if } \tau_i{=}\tau_j = \text{Ia} \\ 0 & \text{otherwise,} \end{cases} \tag{4.12}$$

where $n_{i,j}$ is the bin to which the object belongs. The off-diagonal term, V_{ij}, contains the correlations between supernovae errors that increase in strength with redshift. To produce the step-like structure, $n_{i,j} = \lfloor \frac{min(i,j)}{N/5} \rfloor + 1$ (where "$\lfloor\rfloor$" indicates the floor function, rounding down to the nearest integer). For this covariance matrix we used $s_k = 0.015$ for $k = 1$ to 5, as illustrated in the top left panel of Figure 4.9.

4.4.2.2 Decaying Covariance Matrix

This covariance matrix assumes positive correlations between objects which are nearby in the covariance matrix, with the correlations decaying as the distance between the indices of objects is increased. The exact form of this covariance matrix is

$$C_{ij} = \begin{cases} \sigma^2_{\rm Ia} & \text{if i=j \& } \tau_i{=}\tau_j = \text{Ia} \\ \sigma^2_{\rm nIa} & \text{if i=j \& } \tau_i{=}\tau_j = \text{nIa} \\ \frac{x\sigma^2_{\rm Ia}}{|i-j|} & \text{if i}{\neq}\text{j \& } \tau_i{=}\tau_j = \text{Ia} \\ 0 & \text{otherwise} \end{cases} \tag{4.13}$$

This is illustrated in the middle left panel of Figure 4.9, where we set $x = 0.7$[3].

4.4.2.3 Block-Diagonal Covariance Matrix

To illustrate our method with a more realistic example, we used the covariance matrix for the Union2 sample of 557 supernovae. [42, 126] constructed this matrix by parameterising all known

[3]This covariance matrix is not in general guaranteed to be positive definite. We ensured the realisation of the covariance matrix was positive definite for all analyses done in this work by selecting correct parameter values.

sources of correlations for the Union2 dataset, fitting these nuisance parameters simultaneously with the cosmological parameters and providing an estimate of the best-fit covariance matrix. We binned this into 11 redshift bins, which we then applied to our mock data. An example for one mock data realisation can be seen in the bottom left panel of Figure 4.9. We have set nIa-nIa and nIa-Ia correlations to zero, since these should be very small compared to the intrinsic dispersion of the nIa's. We added a σ_i^2 term to the diagonal, where $\sigma_i = \sigma_{\rm Ia}$ if object i is a Ia and $\sigma_i = \sigma_{\rm nIa}$ if it is a nIa. The block diagonal structure arises largely from the fact that supernovae data from the same survey are correlated with each other.

4.4.3 Numerical Marginalisation Over Supernova Type

4.4.3.1 Theory

A solution to the computational problem of handling correlated data would be to perform the marginalisation over the types numerically instead of analytically, allowing us to use the correlated BEAMS posterior without the 2^N sum. In order to do this, we create N discrete nuisance parameters, the types of the supernovae, τ_i, and marginalise over these parameters in our Markov Chain Monte Carlo (MCMC) (see Section 3.3.1) analysis. This problem is similar to the Ising spin problem because these parameters are discrete and can only assume one of two values [127].

In each step of the MCMC chain, we randomly select one object and set the corresponding τ_i =Ia with probability P_i, which significantly speeds up convergence when compared with varying all the types at each step. To further improve convergence, we choose the initial type parameters based on the ratio of Ia to nIa uncorrelated likelihoods for each object. If the ratio is greater than one, we set the initial type for that object to a Ia, otherwise to a nIa. To compute this initial likelihood ratio, we use some fiducial values for the cosmological parameters. This initial choice for the types has little bearing on the final result, but without it, the chain typically starts with a very low likelihood because many objects have the wrong type and it takes much longer to reach the region of the peak. Thus, if $\boldsymbol{\theta}$ represents the set of parameters in the MCMC chain and $\boldsymbol{\tau}$ the types, the usual MCMC acceptance criterion is

$$R = \min\left(\frac{\mathcal{L}(\boldsymbol{\theta}_{\rm new}, \boldsymbol{\tau}_{\rm new})}{\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\tau})}, 1\right) \tag{4.14}$$

We applied this technique using MCMC to estimate cosmological parameters from the mock datasets. We ran at least three MCMC chains for each dataset produced and ensured all chains were converged using the Gelman-Rubin [84] criterion for convergence (see Section 3.3.1), with $R < 1.01$. Chains were, on average, 60000 steps long after burn-in was removed. Burn-in was determined visually by looking at traceplots of the cosmological parameters as a function of step number (see Figure 4.8): the iteration at which the chain was stationary around a point was considered the end of burn-in. They were found to be uncorrelated above a length of about 30 steps, indicating acceptable mixing, although for MCMC correlations between steps are not in general a problem, if the chain is long enough [128].

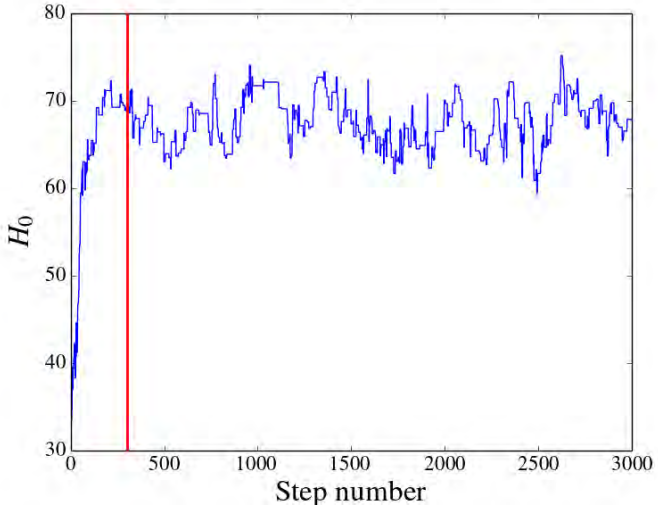


FIGURE 4.8: A subset of a chain indicating with a red line where the burn-in period ends.

It should be noted that for all MCMC chains, the following flat, wide priors were used, to ensure unbiased parameter estimation: $-0.2 < \Omega_m < 1.2$, $-0.2 < \Omega_\Lambda < 1.2$ and $10 < H_0 < 130$. Negative densities are of course, unphysical but we allow such wide priors here due to the small datasets used having poor constraints. An maximum posterior estimate of a negative value would indicate a problem in the analysis. We allowed the shift, b, of the nIa population to vary with a flat prior of $1 < b < 3$ and the standard deviations of the populations, $\sigma_{\rm Ia}$ and $\sigma_{\rm nIa}$ to vary in log space with priors of $-3.0 < \log(\sigma_{\rm Ia}) < -1.2$ and $-0.7 < \log(\sigma_{\rm nIa}) < 1.5$.

To allow for the most general case where only the form of the covariance matrix is known, but its parameters are not, we also varied the covariance parameters for each individual covariance matrix over a flat prior, except for the block covariance matrix whose off-diagonal terms remained fixed, since they were determined from the Union2 dataset.

4.4.3.2 Testing BEAMS with Numerical Marginalisation

We first generated uncorrelated datasets to directly compare the numerical and analytic marginalisation of BEAMS. We found that the $\Omega_m - \Omega_\Lambda$ credible contours[4] produced using Eq.(4.7) and those produced using the numerical marginalisation technique were identical for the uncorrelated datasets, showing that the technique works as expected.

It is interesting to note that the marginalisation method is faster than uncorrelated BEAMS, since there are half the calculations to perform at each step (uncorrelated BEAMS computes the likelihood assuming the object is nIa and Ia respectively, whereas the numerical marginalisation only requires the likelihood given the current types). However, it takes longer for the correlated BEAMS chains to converge. On average, using the Gelman-Rubin [84] criterion for convergence, uncorrelated BEAMS chains converge within 10 000 steps whereas the correlated BEAMS chains only converge after about 40 000 steps. As the dataset scales to larger numbers of supernovae, the *number of steps* taken to converge does not increase appreciably but the likelihood scales roughly as N^2, due to the matrix multiplication between the covariance matrix and the data. However, realistic covariance matrices are likely to be block diagonal, thus we could take advantage of the relative sparsity of the matrix and separate the data to run in parallel. Hence, computational complexity should not be an issue for correlated BEAMS.

Figure 4.9 compares the uncorrelated BEAMS approach with correlated BEAMS (using numerical marginalisation of types) in the case of correlated data, based on the three different covariance matrices from Section 4.4.2. The uncorrelated approach includes no information about correlations and hence it is not surprising that it is biased at $> 2\sigma$ for both the decaying covariance matrix and the wedding cake covariance matrix. In contrast, correlated BEAMS with numerical marginalisation correctly estimates the cosmological parameters without any bias. There is less of an effect from the block diagonal covariance matrix, because both the correlations and the dataset are small. Although even in this case, it is clear the uncorrelated BEAMS contours appear to be mildly biased.

[4] In this paper, we use credible intervals, the Bayesian analogue to confidence intervals. Bayesian credible intervals represent the degree of belief in the parameter estimates, and are derived from the posterior probability. They are not, in general, equivalent to frequentist confidence intervals [77].

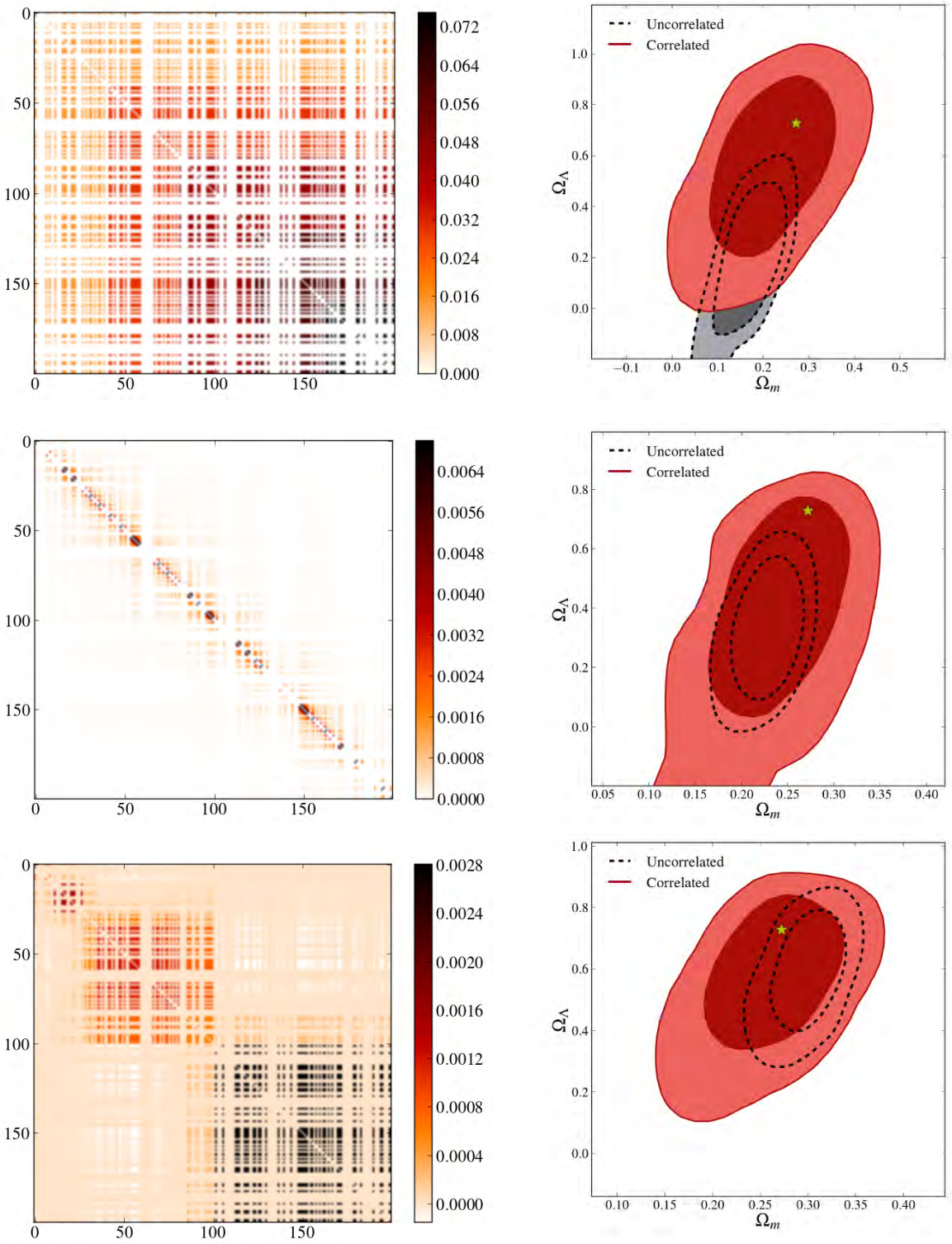


FIGURE 4.9: *Results of the numerical marginalisation of types with BEAMS using three different covariance matrices:* The left column shows schematic views of the three covariance matrices, where the colour indicates the amount of correlation between two data points. The right column shows the $\Omega_m - \Omega_\Lambda$ contours, marginalising over H_0, when the correlated data sets are analysed assuming no correlations (dotted lines) and using correlated BEAMS (filled contours). The covariance matrices and their parameters are (from top to bottom): the wedding cake covariance matrix (s_k=0.015 for k=1 to 5), the decaying covariance matrix (x=0.7) and the block diagonal covariance matrix. All covariance parameters (for the wedding cake and decaying matrices) are marginalised over with wide, flat priors. Note: for all matrices, the diagonal has been removed for clarity because $\sigma_{\rm nIa}$ is so much larger than the off-diagonal terms.

4.4.3.3 Testing the Contours and Coverage Properties of the BEAMS Estimators

To test the accuracy of the correlated BEAMS contours, we created 5000 datasets, each consisting of 200 correlated supernovae, using the wedding cake covariance matrix. We then ran a 100 000 step MCMC chain for both uncorrelated and correlated BEAMS on each of the datasets, totaling about a billion MCMC steps. Figure 4.10 shows scatter plots for both uncorrelated BEAMS and correlated BEAMS. Each point represents the maximum posterior values of the parameters for one of the 5000 datasets. We computed the mean squared error (MSE), which is the sample average of the squared distance between the estimates and the true value for the parameters (in this case Ω_m and Ω_Λ), for both uncorrelated and correlated BEAMS. We found the relative efficiency, defined to be the ratio of the MSE for uncorrelated BEAMS to that of correlated BEAMS, to be 2.3, which implies that correlated BEAMS is a much more efficient estimator than uncorrelated BEAMS for correlated data, i.e. it estimates parameters with much less scatter.

We also plotted the 95% credible contours for five randomly selected datasets to show how the size and shape of the contours are underestimated by uncorrelated BEAMS but well estimated by correlated BEAMS, for these correlated datasets. To quantify this point, we computed the coverage, which we here define to be the proportion of the 5000 datasets where the true value (marked by the black circle in Figure 4.10) lies within the 95% credible interval for each dataset derived from the corresponding MCMC chain for that dataset. We found that the coverage was 88.2% for correlated BEAMS and only 7.2% for uncorrelated BEAMS, showing that accounting for the correlations is crucial in getting the correct contours. Ideally, the coverage should be close to 95%. However, the reader is cautioned that coverage is a frequentist concept and only for asymptotically large datasets would we expect the frequentist coverage to coincide with the corresponding coverage from Bayesian credible intervals. As our datasets only have 200 points each, we should not be surprised if the coverage is not exactly 95%. Undercoverage occurs even for the standard χ^2 method applied to supernovae datasets (see Figure 8 in [129]).

The coverage properties of both estimators are illustrated in Figure 4.11. For both uncorrelated and correlated BEAMS we plotted the effective coverage (on the y-axis), i.e. the fraction of datasets for which the given credible contour contains the input cosmological parameters, at that level of credibility (on the x-axis). So for example, the 0.95 credible level corresponds to a coverage of 0.882 and 0.072 for correlated and uncorrelated BEAMS respectively. Ideal coverage is the diagonal straight line across the plot, where (for example) 0.95 of the datasets

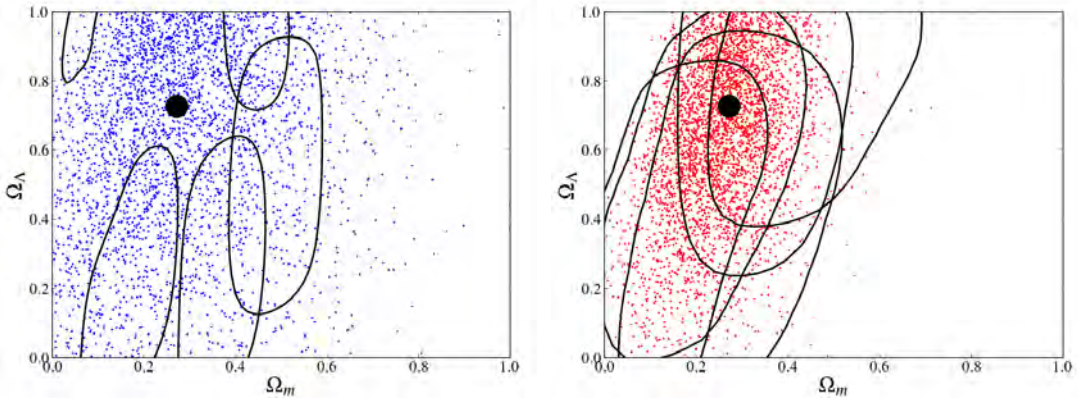


FIGURE 4.10: Scatter plots of the maximum posterior estimates for the cosmological parameters for each of 5000 correlated datasets using the wedding cake covariance matrix (Eq.(4.11)). *Left panel:* BEAMS assuming the data are uncorrelated. *Right panel:* correlated BEAMS with numerical marginalisation over covariance matrix parameters and types. Also plotted are the 95% credible contours from five randomly selected datasets to give an idea of the size and shape of the contours. This shows that, while both methods are unbiased on average, correlated BEAMS is a much better estimator and has far superior coverage properties (88.2% vs. only 7.2%).

would contain the input cosmology in the 0.95 credible contour. It is clear that the coverage of correlated BEAMS is close to ideal and far superior to that of uncorrelated BEAMS.

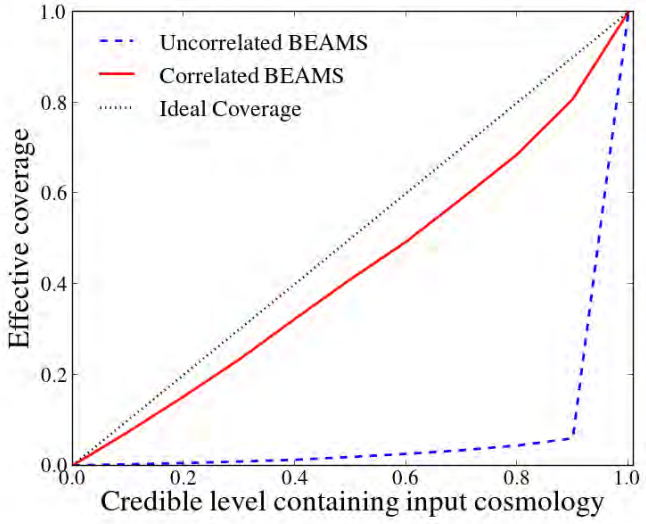


FIGURE 4.11: Effective coverage, defined to be the fraction of datasets which contain the input cosmology at a given credible level for the 5000 datasets shown in Figure 4.10 for both uncorrelated BEAMS (blue, dashed line), correlated BEAMS (red, solid line) and for the ideal coverage case (black, dotted line). While correlated BEAMS is close to ideal, implying that its contours can be trusted, uncorrelated BEAMS dramatically undercovers and hence the contours cannot be trusted for correlated data, as expected.

4.5 A Perturbative Expansion of the BEAMS Posterior

For future surveys, in exactly the limit where it will be important to apply BEAMS to avoid biases from probability cuts, the light curve data for both candidates and templates will be excellent and the probability that most candidates are a Ia will therefore be close to zero or unity. This is highlighted in Figure 4.7, which shows simulated DES probabilities using MLCS2k2 [96] and SNANA [104]. In this limit of abundant, high-quality data, we may expect we could perform a perturbative expansion of the full correlated BEAMS posterior, to find a computationally feasible approximation by reducing the number of terms.

4.5.1 Theory

We start with the BEAMS posterior in the general case:

$$P(\theta|D) = \sum_{\tau} P(D|\theta, \vec{\tau})P(\theta)P(\vec{\tau})/P(D) \tag{4.15}$$

This results in 2^N terms being calculated, in the case where there are two types.

We can Taylor expand this posterior, written as a function of the probabilities $\vec{p}$ (this is a length-N vector containing the probability for each object), around the point where the probabilities are rounded to either zero or one. We define the vector $\vec{\epsilon}$ such that $\epsilon_i = p_i - Rnd(p_i)$, where $Rnd(p_i)$ is the rounded value of the probability of the i'th data point (which will be either 0 or 1). So for example, if $p_i = 0.99$ then $\epsilon_i = -0.01$ and if $p_i = 0.02$ then $\epsilon_i = 0.02$ etc.

The BEAMS posterior becomes:

$$P(\theta|D,\vec{p}) \propto P(\theta|D,\vec{p})\Big|_{P_{Rnd}} + \vec{\epsilon}\cdot\nabla P(\theta|D,\vec{p})\Big|_{P_{Rnd}} + \frac{1}{2}\vec{\epsilon}\cdot\left[\vec{\epsilon}\cdot H(P(\theta|D,\vec{p}))\right]\Big|_{P_{Rnd}} + \cdots, \tag{4.16}$$

where H is the Hessian matrix.

4.5.1.1 The First Order Term

Now we will explicitly calculate the first order derivative term. The i'th component of $\nabla P(\theta|D,\vec{p})$ is given by

$$\frac{\partial P(\theta|D,\vec{p})}{\partial p_i} = \sum_{\tau} P(D|\theta,\vec{\tau})P(\theta)\frac{\partial P(\vec{\tau})}{\partial p_i}. \tag{4.17}$$

This derivative is trivial since $P(\vec{\tau})$ is linear in the p_i's. $P(\vec{\tau})$ is defined as

$$P(\vec{\tau}) = \prod_{\tau_j=\mathrm{Ia}} p_j \prod_{\tau_k=\mathrm{nIa}} (1-p_k)\,. \tag{4.18}$$

Since only one of these terms is a function of p_i (where i equals either j or k depending on its type), the derivative is given by

$$\frac{\partial P(\vec{\tau})}{\partial p_i} = \frac{\partial P_{\tau_i}}{\partial p_i} P(\vec{\tau}_{-i}). \tag{4.19}$$

Here, $P(\vec{\tau}_{-i})$ is equivalent to Eq.(4.18), with the contribution to the product from the i'th object's probability removed such that $P(\vec{\tau}) = P_{\tau_i}P(\vec{\tau}_{-i})$. Since P_{τ_i} is either p_i or $(1-p_i)$, its derivative is given by

$$\frac{\partial P_{\tau_i}}{\partial p_i} = \begin{cases} 1 & \text{if } \tau_i = \mathrm{Ia} \\ -1 & \text{if } \tau_i = \mathrm{nIa}. \end{cases} \tag{4.20}$$

Thus:

$$\frac{\partial P(\theta|D,\vec{p})}{\partial p_i} = \sum_{\tau} \pm P(D|\theta,\vec{\tau})P(\theta)P(\vec{\tau}_{-i})\,. \tag{4.21}$$

Once we subsitute for the rounded probabilities, many of the terms in $P(\vec{\tau}_{-i})$ will go to zero. In fact, each element of $\vec{\tau}_{-i}$ must equal its rounded type for $P(\vec{\tau}_{-i})$ in order to be non-zero, in which case, $P(\vec{\tau}_{-i}) = 1$. This leaves only two terms, corresponding to the two possible types of the i'th data point. Thus:

$$\left.\frac{\partial P(\theta|D,\vec{p})}{\partial p_i}\right|_{P_{Rnd}} = \left[P(D|\theta,\vec{\tau}_{i=\mathrm{Ia}}) - P(D|\theta,\vec{\tau}_{i=\mathrm{nIa}})\right] P(\theta)\,, \tag{4.22}$$

where $\vec{\tau}_{i=\mathrm{Ia}}$ is the same as $\vec{\tau}_{Rnd}$ (where $\tau_{Rnd_i} = \mathrm{Ia}$ if $Rnd(p_i) = 1$ and $\tau_{Rnd_i} = \mathrm{nIa}$ if $Rnd(p_i) = 0$) with the i'th element being Type Ia and $\vec{\tau}_{i=\mathrm{nIa}}$ is the same except the i'th element is Type nIa.

Finally, the first order BEAMS posterior is given by

$$P(\theta|D,\vec{\epsilon}) \propto P(\theta|D,\vec{p})\Big|_{P_{Rnd}} + \sum_{i=1}^{N}\left(\epsilon_i \times [P(D|\theta,\vec{\tau}_{i=\mathrm{Ia}}) - P(D|\theta,\vec{\tau}_{i=\mathrm{nIa}})] \times P(\theta)\right). \quad (4.23)$$

For N data points, the order of this calculation is $2N$ times the order of the likelihood calculation. So in the case where the likelihood is given by the usual Gaussian form $e^{-\chi^2/2}$, this would be an order $2N^2$ calculation.

4.5.1.2 The Second Order Term

Our treatment can be repeated for the second order term. The second order derivative is given by

$$\frac{\partial^2 P(\theta|D,\vec{p})}{\partial p_i \partial p_j} = \sum_{\tau} P(D|\theta,\vec{\tau})P(\theta)\frac{\partial^2 P(\vec{\tau})}{\partial p_i \partial p_j}\,. \quad (4.24)$$

$P(\vec{\tau})$ can again be expressed as a product of one i-dependent term, one j-dependent term and one term independent of both

$$P(\vec{\tau}) = P_{\tau_i} P_{\tau_j} P_{\tau_{-i,j}}\,, \quad (4.25)$$

if $i \neq j$. If $i = j$, then the second derivative of $P(\vec{\tau})$ will be zero.

By the product rule, this is

$$\frac{\partial^2 P(\vec{\tau})}{\partial p_i \partial p_j} = \frac{\partial P_{\tau_i}}{\partial p_i}\frac{\partial P_{\tau_j}}{\partial p_j} P(\vec{\tau}_{-i,j})\,. \quad (4.26)$$

Which evaluates to

$$\frac{\partial^2 P(\vec{\tau})}{\partial p_i \partial p_j} = \begin{cases} 0 & \text{if i=j} \\ 1 & \text{if } i \neq j \text{ and } \tau_i = \tau_j \\ -1 & \text{if } i \neq j \text{ and } \tau_i \neq \tau_j. \end{cases} \quad (4.27)$$

Similarly to the first order term, when we evaluate the full posterior (Eq.(4.24)) at the rounded probabilities, we find that only four terms remain (still for $i \neq j$):

$$\left(\frac{\partial^2 P(\theta|D,\vec{p})}{\partial p_i \partial p_j}\right)\Big|_{P_{Rnd}} = \Big[P(D|\theta,\vec{\tau}_{i,j=\mathrm{Ia}}) + P(D|\theta,\vec{\tau}_{i,j=\mathrm{nIa}}) - \\ -P(D|\theta,\vec{\tau}_{i=\mathrm{Ia,j=nIa}}) - P(D|\theta,\vec{\tau}_{i=\mathrm{nIa,j=Ia}})\Big]\Big[P(\theta)\Big], \quad (4.28)$$

where

$$\begin{aligned}
\vec{\tau}_{i,j=\mathrm{Ia}} &= \vec{\tau}_{Rnd} \quad \text{except with} \quad \tau_i = \tau_j = \mathrm{Ia} \\
\vec{\tau}_{i,j=\mathrm{nIa}} &= \vec{\tau}_{Rnd} \quad \text{except with} \quad \tau_i = \mathrm{Ia}, \tau_\mathrm{j} = \mathrm{nIa} \\
\vec{\tau}_{i=\mathrm{Ia,j=nIa}} &= \vec{\tau}_{Rnd} \quad \text{except with} \quad \tau_i = \mathrm{nIa}, \tau_\mathrm{j} = \mathrm{Ia} \\
\vec{\tau}_{i=\mathrm{nIa,j=Ia}} &= \vec{\tau}_{Rnd} \quad \text{except with} \quad \tau_i = \tau_j = \mathrm{nIa}.
\end{aligned}$$

Putting this all together, the second order term of the BEAMS posterior is

$$\frac{1}{2}\vec{\epsilon}\cdot\Big[\vec{\epsilon}\cdot H(P(\theta|D,\vec{p}))\Big]\Big|_{P_{Rnd}} = \frac{1}{2}\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N}\Bigg(\epsilon_i \times \epsilon_j \times \Big[P(D|\theta,\vec{\tau}_{i,j=\mathrm{Ia}}) + P(D|\theta,\vec{\tau}_{i,j=\mathrm{nIa}}) - \\ -P(D|\theta,\vec{\tau}_{i=\mathrm{Ia,j=nIa}}) - P(D|\theta,\vec{\tau}_{i=\mathrm{nIa,j=Ia}})\Big]\Big[P(\theta)\Big]\Bigg). \quad (4.29)$$

Thus the second order term is order $(4N)^2$ times the order of the likelihood calculation. For a Gaussian likelihood, this is an order N^3 calculation.

4.5.2 Results

This perturbative approximation breaks down because we expand the posterior about the rounded off probabilities (that is, if the probability is close to 1 we take the object to be a Ia and a nIa if it is close to 0), but due to the extreme nature of the likelihood, this tends to lie quite far from the maximum likelihood, the point about which one would like to perform the expansion. This happens because $\sigma_{\mathrm{Ia}}/\sigma_{\mathrm{nIa}}$ is very small, hence if too many nIa's are mistyped as Ia's when rounding off the probabilities, these terms cause higher order terms in the Taylor expansion to dominate, however we only include terms up to the second order term in this expansion in order for the analysis to remain computationally viable (the second order term is

an N^3 calculation, the third order N^4 and so on). Thus the perturbative expansion is not a good enough approximation of the true likelihood.

This can be seen in Figures 4.12 and 4.13, where we show 1 and 2-σ contours in the $\Omega_m - \Omega_\Lambda$ plane. In each case, the fiducial value is shown by a star. While the underlying data are correlated (using the decaying covariance matrix model), the uncorrelated form of BEAMS *assumes* uncorrelated data. In addition, perturbative BEAMS cannot sufficiently correct for the mistyped terms, and thus fail to recover the simulated data model at $> 3\sigma$.

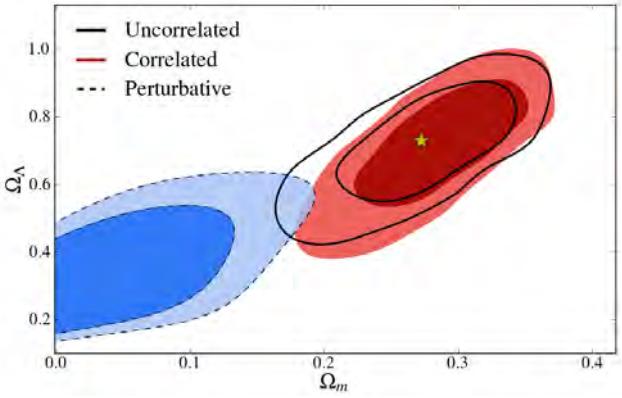


FIGURE 4.12: Comparison between uncorrelated BEAMS (line contours), correlated BEAMS (filled, red contours) and the perturbative expansion of BEAMS (filled blue contours with dashed lines), for an uncorrelated dataset, showing how the perturbative expansion fails badly even in the uncorrelated case.

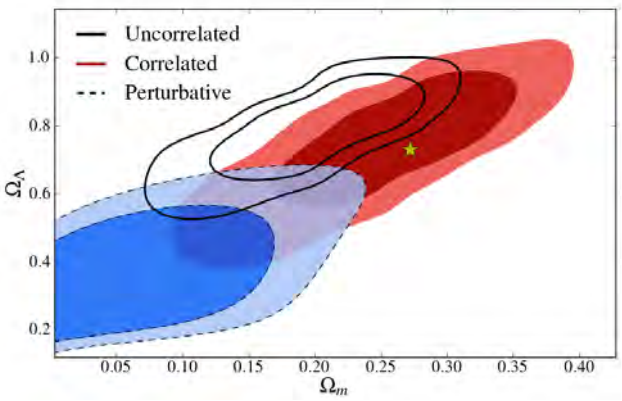


FIGURE 4.13: Comparison between uncorrelated BEAMS (line contours), correlated BEAMS (filled, red contours) and the perturbative expansion of BEAMS (filled blue contours with dashed lines), for a correlated dataset, using the decaying diagonal covariance matrix, showing how the perturbative expansion fails badly.

The failure of perturbative BEAMS to correct for mistyped terms can be understood further by considering Figure 4.14. In the left panel, we show perturbative BEAMS successfully approximates the true likelihood, as does correlated BEAMS using numerical marginalisation,

for a small dataset. The right panel, however, shows this is not the case for a large dataset. The perturbative BEAMS estimate for Ω_Λ is biased, because mistyped objects in the dataset cause higher order terms in the perturbative approximation to dominate. As shown previously, however, correlated BEAMS (with numerical marginalisation of types) accurately recovers the input cosmology.

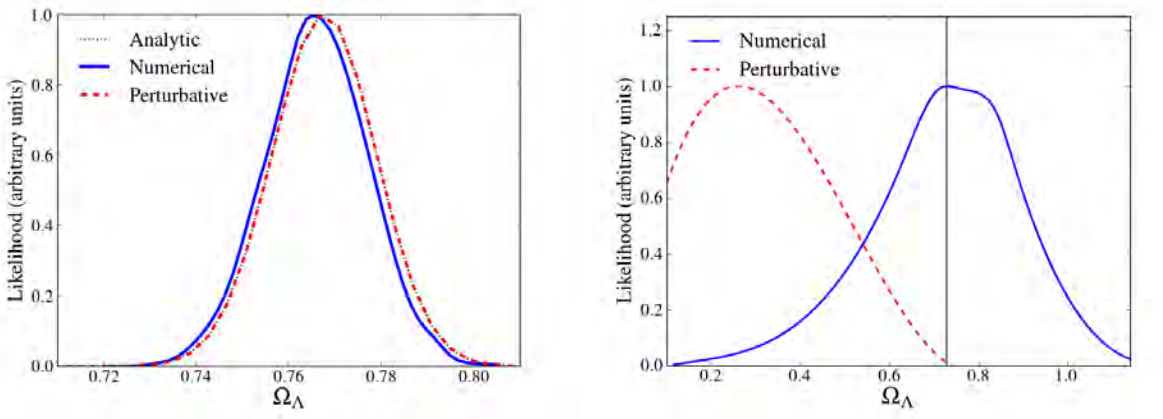


FIGURE 4.14: These plots demonstrate the difference between the true BEAMS likelihood (dotted, black line), the perturbative approximation (dashed, red line) and correlated BEAMS with numerical marginalisation (solid, blue line). *Left panel:* A set of ten data points was created using the decaying covariance matrix to correlate the data. Ω_m and H_0 are set to 0.272 and 70.4 respectively and Ω_Λ is allowed to vary. For this small dataset, perturbative BEAMS and correlated BEAMS are both good approximations to the true BEAMS likelihood. *Right panel:* A similar dataset except with 200 data points this time. Now, perturbative BEAMS poorly approximates the true likelihood and produces a bias in Ω_Λ. BEAMS with numerical marginalisation, however, is unbiased. The true likelihood cannot be calculated with such a large dataset.

4.6 Discussion

Photometric supernova surveys with unprecedented amounts of data will provide an exciting opportunity to learn about the structure and evolution of the universe. Due to the vast number of supernova candidates and their extended redshift distribution in large future surveys, it will be difficult, if not impossible, to spectroscopically follow-up all candidates as is normally required in cosmological analyses. BEAMS is a rigorous statistical method which avoids biases while using all supernova candidates, together with the probability that a candidate is a Type Ia supernova, derived from the multicolour lightcurves of the candidate. Until now, BEAMS has been applied assuming the supernovae data are uncorrelated [45–47], an assumption which will be inappropriate for future surveys. Without this assumption, the analytical form of the BEAMS posterior is computationally unfeasible. If the uncorrelated form of BEAMS is applied to a

dataset with correlated systematic uncertainties, the posterior for the cosmological parameters can be incorrectly estimated.

To deal with this 'exponential catastrophe', we have explored two different approaches. The first marginalises over all the possible combinations of object type numerically instead of analytically, by including the types as discrete nuisance parameters in our MCMC chains, making it computationally efficient. We have shown, with three separate models of covariance matrices, that this algorithm successfully recovers the input cosmology in the correlated case without bias. In addition, we have shown with 5000 mock datasets that the correlated BEAMS credible contours are reliable estimates of the true error contours. This is something that cannot be easily reproduced without using the correlated BEAMS formalism.

The second approach we considered was a perturbative expansion of the BEAMS posterior which typically fails because, when too many objects are mistyped, the higher order terms of the expansion (which are neglected due to computational constraints) dominate over the lower order terms, even when the probabilities are very close to zero or one thus producing a biased posterior. However, with numerical marginalisation over types, correlations between supernovae do not appear to be an impediment to using BEAMS in analysing future photometric supernova surveys.

As supernovae surveys get larger, contamination will become more of a problem, thus increasing the need for a technique such as BEAMS. The Bayesian approach also allows a very natural way to model additional complexities and effects while accurately propagating uncertainties. For example, a future extension to BEAMS would be to deal with photometric redshifts, where each object's redshift would become a parameter to marginalise over. The BEAMS framework lends itself to the incorporation of such extensions. It is also a highly versatile technique which may find application in any field where one wishes to estimate parameters from a dataset that contains contaminants.

Chapter 5

Bayesian Inference for Radio Observations

5.1 Introduction

The current approach to doing science with interferometric radio observations is an exercise in data massaging and manipulation. Because interferometric images have missing data (due to the gaps between telescopes), technically there is an infinite possible number of sky models which could fit the available data. To solve this problem of deconvolution (that is converting the raw visibilities to an image of the true sky) the algorithm of CLEAN [130] has been proposed, amongst others.

CLEAN simply and elegantly extracts the most likely radio sources in the data while removing probable artifacts. The resulting image is then used for science such as estimating the fluxes, shapes and positions of the sources in the image. The problem is, there is no guarantee the image produced is an accurate representation of the sky. Firstly, the CLEAN algorithm requires fairly arbitrary user choices such as stopping criteria, loop gain and in some cases search windows, which can significantly affect the resultant image. Secondly, natural randomness means that the best fitting parameters (such as the fluxes) may lie away from the true parameters (see the histograms in Section 5.6) and, since CLEAN cannot provide accurate uncertainty information [131], there seems to be no rigorous way of knowing just how much one should believe the fluxes extracted from a CLEANed image. Thirdly, the fundamental assumption required by CLEAN that the map can be described as a collection of point sources makes recovering

extended emission with it difficult [132]. Lastly, instrumental errors, such as beam parameters and pointing errors (as we will discuss) and scientific parameters, such as fluxes and extended emission parameters, are often correlated. If the instrumental errors are ignored, or determined (by some other method since CLEAN cannot extract them from the data) and fixed to potentially incorrect values, this will cause biases in the estimated scientific parameters, as we will show.

We propose an alternative approach, doing a full Bayesian analysis from the raw visibilities to the science and instrumental parameters, along the way learning about their relationships and uncertainties, and obtaining the full posterior for the problem (see Chapter 3 for a review of Bayesian statistics). We call this technique Bayesian Inference for Radio Observations or BIRO.

5.2 Current Approaches to Interferometric Radio Observations

CLEAN is the probably the most successful and most popular method to analyse radio interferometric images. The problem CLEAN attempts to solve is that, by its very nature, an interferometer incompletely samples the (u, v) plane (essentially because it is impossible to have a continuum of baselines) and thus, a radio image has missing information. The "dirty beam" of the interferometer may contain artifacts such as grating rings and sidelobes, which will appear in the resulting "dirty map". CLEAN operates on the premise that it is unlikely that the true sky consists of these rather artificial looking shapes. It first makes the fundamental assumption that the map can be described as a collection of point sources. CLEAN proceeds by identifying the strongest source in the map, removing the beam pattern it produces (across the whole map) and then iterating this step until the entire map is consistent with noise. The "cleaned map" then consists of the series of clean components removed by the algorithm, with their correct amplitude and position. CLEAN has been so successful that it has remained the standard for use in astronomy for about forty years, despite its difficulty in reproducing areas of extended emission (although this is helped by multiscale CLEAN [133]) and its inability to determine uncertainties [131]. Figure 5.1 illustrates graphically the CLEAN algorithm by way of a simple example and Figure 5.2 demonstrates CLEAN on a simulated source.

The maximum entropy method (MEM) was used successfully in various other fields such as geosciences and mathematical statistics before Ables [134] and others [135, 136] applied it to radio astronomy. It has remained the main competitor to CLEAN and other related methods.

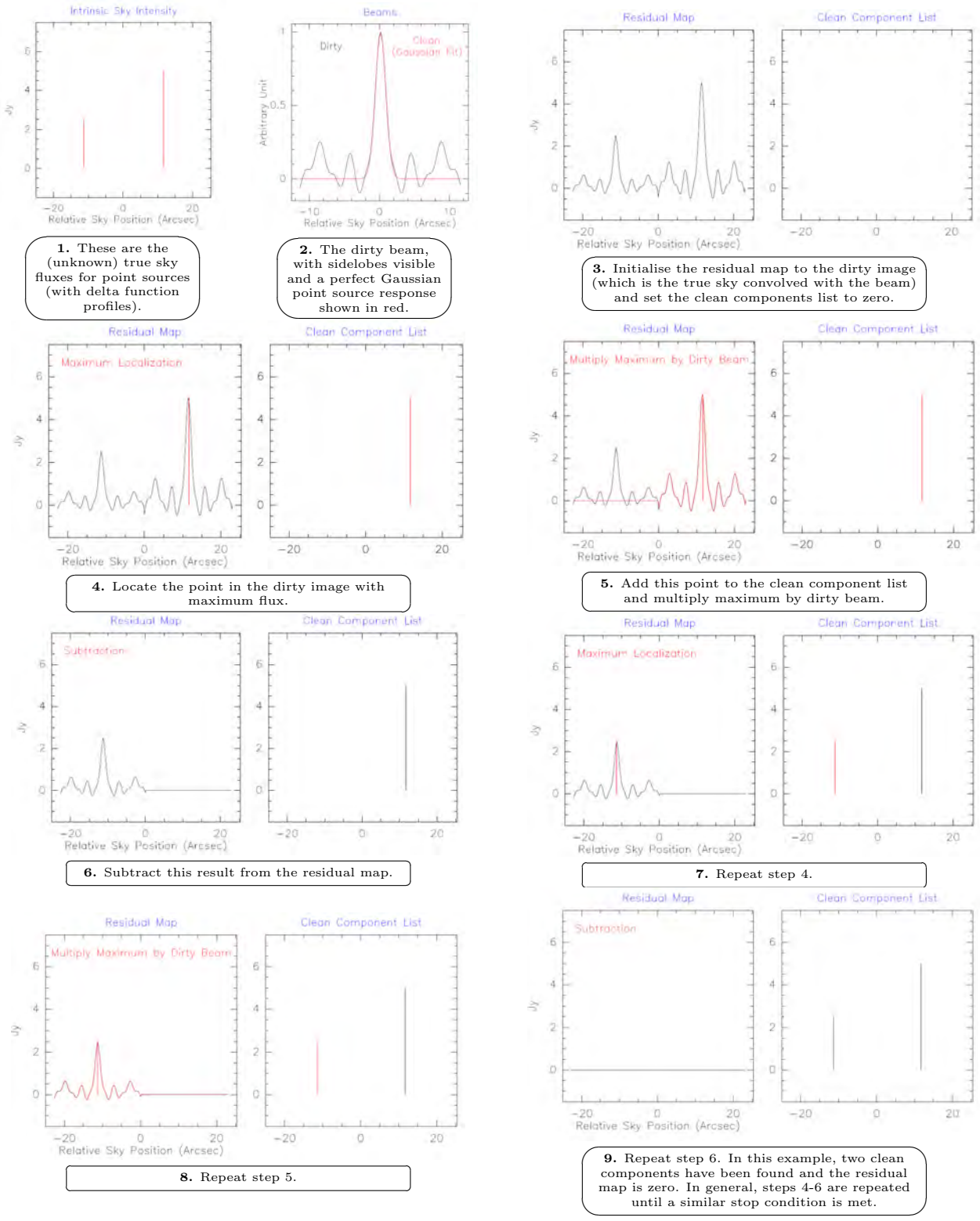


FIGURE 5.1: A demonstration of the CLEAN algorithm for two sources in a dirty image (adapted from [70]).

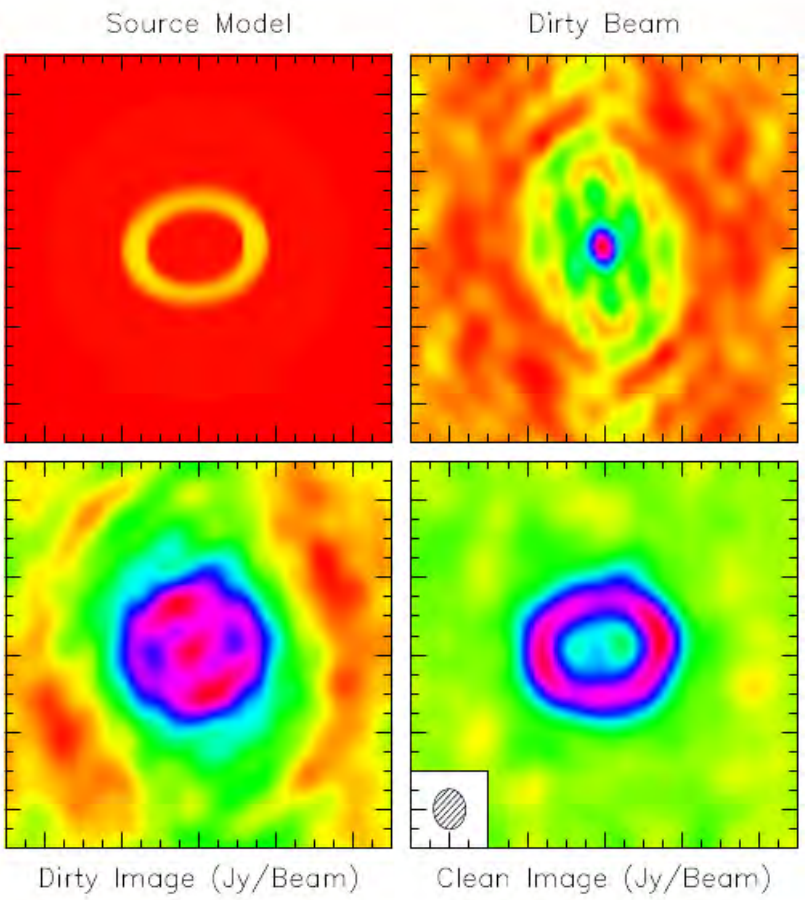


FIGURE 5.2: CLEAN applied to a simulated source. The source (top left) is convolved with the dirty beam (top right) to produce the dirty map (bottom left). This is then CLEANed following the algorithm outlined in Figure 5.1 to produce the clean image (bottom right) [70].

The MEM works by producing an image which has the smallest amount of structure (hence the maximum entropy) whilst still being consistent with original observations. Sutton *et al.* [132] extend the MEM by producing maps which maximise the posterior probability density (related to the entropy), given a model of the sky.

RESOLVE (Radio Extended SOurces Lognormal deconVolution Estimator) [131, 137] is a similar algorithm which finds the optimal sky by maximising the *a posteriori* probability density of the map, with the addition that the uncertainty of the distribution can be approximated, and which is optimised for dealing with extended emission.

UVMULTIFIT [138] is a useful tool to consider in this context because it has been developed again as an alternative to CLEAN, to allow the user to fit sky models directly to their visibility data. The authors argue that analysing interferometric images after deconvolution is like "fitting

models to models", whereas one would rather fit models directly to data. A similar approach is applied to photon counts in [139], where a hierarchical Bayesian model is applied.

Schammel *et al.* [140] and Carvalho *et al.* [141, 142] have implemented the Bayesian techniques similar to those of Hobson & McLachlan [13], to detect astronomical objects in diffuse backgrounds. These techniques can use model selection to determine the most likely number of sources and are also capable of marginalising over nuisance parameters, although they have computational limitations related to the number of parameters involved.

A joint Bayesian analysis has successfully been applied several times to AMI (the Arcminute Microkelvin Imager) [143] data in order to detect Sunyaev-Zel'dovich (SZ) clusters [144, 145] at the same time as constraining instrumental nuisance parameters. Similarly, Lancaster *et al.* [146] performed a joint Bayesian analysis for the detection of SZ clusters in the presence of contaminating CMB noise and foreground sources.

Clearly, there is a precedent for this type of Bayesian analysis in specific cases. What we propose is a general framework which will be applicable for any radio interferometric observation (although in principle could be extended to any wavelength observation where the sky could be reasonably well modeled), and could be used even for as large an instrument as the SKA[1], where millions of parameters may have to be fit simultaneously.

5.3 The Radio Interferometry Measurement Equation

The Radio Interferometry Measurement Equation (RIME) originally introduced by Hamaker *et al.* [72] provides a tidy, rigorous, mathematical approach to radio interferometry modeling and calibration. Before the introduction of the RIME, approaches to this problem were somewhat *ad hoc*, lacking a full mathematical backing, and approximate [71]. The RIME allows complete modeling in a very simple formalism, which allows one to include all known instrumental and observational effects that interfere with the signal before arriving at the telescope. We utilise the RIME heavily in the development of our Bayesian approach to radio observations.

[1] Square Kilometre Array, http://www.skatelescope.org

5.3.1 Derivation of the RIME

What follows here is the derivation of the RIME from Smirnov (2011) [71], which is a rederivation of the original formalism [72] under a more consistent mathematical framework.

We first start with a single point source (quasi-monochromatic), the signal of which can be described by the complex vector $\boldsymbol{e}$ that represents the strength of the electric field. The intensity of Eq.(2.6) is related to the electric field as described in Eq.(5.10), such that $I_\nu \propto |\boldsymbol{e}|^2$. In a coordinate system where the z axis is aligned with the direction of propagation, $\boldsymbol{e}$ can be represented as a column vector of two complex numbers:

$$\boldsymbol{e} = \begin{pmatrix} e_x \\ e_y \end{pmatrix}. \tag{5.1}$$

Now as the signal passes from the source to the receiver, it gets affected by multiple observational effects (including ionospheric and atmospheric effects) as well as by instrumental effects. If we assume these transformations are linear, the resulting signal will be:

$$\boldsymbol{e'} = \boldsymbol{J}\boldsymbol{e}, \tag{5.2}$$

where $\boldsymbol{J}$ is the 2x2 complex matrix known as the Jones matrix[74]. The Jones matrix is a concise way to include the previously mentioned effects. Shortly we will give some example Jones matrices.

Each effect in the signal path can be represented by a Jones matrix and as each effect is applied to the signal, its Jones matrix is multiplied, resulting in a *Jones chain*:

$$\boldsymbol{e'} = \boldsymbol{J}_n \boldsymbol{J}_{n-1} \dots \boldsymbol{J}_1 \boldsymbol{e} = \boldsymbol{J}\boldsymbol{e}. \tag{5.3}$$

Note that the order in which the Jones matrices are applied is important because it corresponds to the order of effects along the signal path and matrix multiplication does not, in general, commute.

Converting the electromagnetic signal that arrives at the telescope to voltages which are ultimately recorded involves several instrumental effects which can, of course, be described by Jones matrices. Let us assume we have two antenna feeds a and b (which may be, for example, two linear dipoles or left and right circular feeds). Then we can write the voltage as vector, $\boldsymbol{v}$

equal to the total Jones matrix, $\boldsymbol{J}$, times the original signal, $\boldsymbol{e}$:

$$\boldsymbol{v} = \begin{pmatrix} v_a \\ v_b \end{pmatrix} = \boldsymbol{J}\boldsymbol{e}. \tag{5.4}$$

Now to apply this to an interferometer, let us consider two antennae, p and q, which measure two independent voltage vectors, $\boldsymbol{v}_p$, $\boldsymbol{v}_q$ of the same source. To retrieve a coherent signal from these two antennae, the correlator performs pairwise correlations (see Eq.(2.8) and Eq.(2.9)) between the components of $\boldsymbol{v}_p$ and $\boldsymbol{v}_q$:

$$\langle v_{pa} v_{qa}^* \rangle, \langle v_{pa} v_{qb}^* \rangle, \langle v_{pb} v_{qa}^* \rangle, \langle v_{pb} v_{qb}^* \rangle, \tag{5.5}$$

where $\langle\rangle$ represents averaging over some small time and frequency bin and * is the complex conjugate. Let us arrange these correlations into the visibility matrix, V_{pq}:

$$\mathsf{V}_{pq} = 2 \begin{pmatrix} \langle v_{pa} v_{qa}^* \rangle & \langle v_{pa} v_{qb}^* \rangle \\ \langle v_{pb} v_{qa}^* \rangle & \langle v_{pb} v_{qb}^* \rangle \end{pmatrix} [2]. \tag{5.6}$$

Then V_{pq} can be written as the following matrix product:

$$\mathsf{V}_{pq} = 2 \left\langle \begin{pmatrix} v_{pa} \\ v_{pb} \end{pmatrix} (v_{qa}^*, v_{qb}^*) \right\rangle = 2 \langle \boldsymbol{v}_p \boldsymbol{v}_q^H \rangle, \tag{5.7}$$

where H is the conjugate transpose operation, also known as the Hermitian transpose.

Now although the signal starts with the same vector, $\boldsymbol{e}$, it follows different paths to the two antennae p and q, each path with its own Jones matrix, $\boldsymbol{J}_p$ and $\boldsymbol{J}_q$. If we combine Eqs. (5.4) and (5.7) we get:

$$\mathsf{V}_{pq} = 2 \langle \boldsymbol{J}_p \boldsymbol{e} (\boldsymbol{J}_q \boldsymbol{e})^H \rangle = 2 \langle \boldsymbol{J}_p (\boldsymbol{e}\boldsymbol{e}^H) \boldsymbol{J}_q^H \rangle. \tag{5.8}$$

Assuming the Jones matrices for the antennae are constant over the averaging interval, they can be moved outside the averaging operator to obtain:

$$\mathsf{V}_{pq} = 2 \boldsymbol{J}_p \langle \boldsymbol{e}\boldsymbol{e}^H \rangle \boldsymbol{J}_q^H = 2 \boldsymbol{J}_p \begin{pmatrix} \langle e_x e_x^* \rangle & \langle e_x e_y^* \rangle \\ \langle e_y e_x^* \rangle & \langle e_y e_y^* \rangle \end{pmatrix} \boldsymbol{J}_q^H. \tag{5.9}$$

[2]The factor of two here is purely convention and is discussed at length in [71]. It is essentially placed here to ensure (amongst other reasons) that the brightness matrix is unity for a 1 Jy unpolarised source and that the "onion form" of the RIME does not require a factor of two to be carried around.

It turns out these bracketed quantities are related to the definition of the Stokes parameters [67, 147] and [73] show that:

$$2\begin{pmatrix}\langle e_x e_x^*\rangle & \langle e_x e_y^*\rangle \\ \langle e_y e_x^*\rangle & \langle e_y e_y^*\rangle\end{pmatrix} = \begin{pmatrix} I+Q & U+iV \\ U-iV & I-Q\end{pmatrix} = \mathsf{B}. \tag{5.10}$$

This defines B, which is called the *brightness matrix* and also leads us to the elegant, condensed form of the RIME for a single point source:

$$\mathsf{V}_{pq} = \boldsymbol{J}_p \mathsf{B} \boldsymbol{J}_q^H. \tag{5.11}$$

As the signal travels from the source to the receivers of the telescopes and ultimately gets converted into voltages, more and more effects or layers of corruptions are added to it, in the form of Jones matrices. This gives rise to the "onion" form of the RIME:

$$\mathsf{V}_{pq} = \boldsymbol{J}_{pn}(\ldots(\boldsymbol{J}_{p2}(\boldsymbol{J}_{p1}\mathsf{B}\boldsymbol{J}_{q1}^H)\boldsymbol{J}_{q2}^H)\ldots)\boldsymbol{J}_{qm}^H. \tag{5.12}$$

Note that as the two signal paths may be quite different, this form is asymmetric.

5.3.2 Applying the RIME

The Jones matrix is an essential aspect of the RIME as it encodes everything that can corrupt and affect a signal as it passes to the receiver. For example, the Jones matrix for phase delay, that is the difference in signal arrival times in the two antennae (see Eq.(2.10)), is a simple scalar matrix of the form:

$$K = \mathrm{e}^{i\phi} \equiv \mathrm{e}^{i\phi}\begin{pmatrix}1 & 0 \\ 0 & 1\end{pmatrix}. \tag{5.13}$$

Now, adopting the conventional coordinate system [67] with the z axis pointing to the phase centre of the telescope, antenna p is located at coordinates $\boldsymbol{u}_p = (u_p, v_p, w_p)$. The phase difference at $\boldsymbol{u}_p$ for a signal arriving from direction $\boldsymbol{\sigma}$ is:

$$\phi_p = 2\pi\lambda^{-1}(u_p l + v_p m + w_p(n-1)), \tag{5.14}$$

where m,l,$n = \sqrt{1-l^2-m^2}$ are the direction cosines of $\boldsymbol{\sigma}$. Expressing $\boldsymbol{u}$ in units of wavelengths, we can now write the scalar K-Jones matrix:

$$K_p = \mathrm{e}^{-i\phi_p} = \mathrm{e}^{-2\pi i u_p l + v_p m + w_p(n-1)}. \tag{5.15}$$

Thus the RIME for a single, uncorrupted point source (which is the ideal case) would be:

$$\mathsf{V}_{pq} = K_p \mathsf{B} K_q^H. \tag{5.16}$$

Or, substituting in Eq.(5.15):

$$\mathsf{V}_{pq} = \mathsf{B}\mathrm{e}^{-2\pi i u_{pq} l + v_{pq} m + w_{pq}(n-1)}, \tag{5.17}$$

where $\boldsymbol{u}_{pq} = \boldsymbol{u}_p - \boldsymbol{u}_q$. This is the visibility matrix expressed in terms of baseline uvw coordinates, $\boldsymbol{u}_{pq}$. We call this form of the visibility matrix, with the phase delay included, the source coherency and write it as X_{pq}.

In reality, there are many other corrupting effects on the signal. We can write the antenna gains for simple linear dipoles as:

$$\boldsymbol{G} = \begin{pmatrix} g_x & 0 \\ 0 & g_y \end{pmatrix}. \tag{5.18}$$

Combining these Jones matrices, we can now write the measurement equation for a single corrupted point source (remembering that scalar matrices commute) as:

$$\mathsf{V}_{pq} = \boldsymbol{G}_p K_p \mathsf{B} K_q^H \boldsymbol{G}_q^H. \tag{5.19}$$

Then we can write down the general RIME for a collection of point sources:

$$\mathsf{V}_{pq} = \boldsymbol{G}_p \left(\sum_s \boldsymbol{E}_{sp} \mathsf{X}_{spq} \boldsymbol{E}_{sq}^H \right) \boldsymbol{G}_q^H, \tag{5.20}$$

where $\boldsymbol{E}_{sp}$ describes the direction dependent effects for each source and $\boldsymbol{G}_p$ describes the direction independent effects.

The full-sky RIME is obtained by integrating as a function of direction [71]:

$$\mathsf{V}_{pq} = \boldsymbol{G}_p \left(\int\int_{lm} \mathsf{B}_{pq} \mathrm{e}^{-2\pi i (u_{pq} l + v_{pq} m)} \mathrm{d}l \mathrm{d}m \right) \boldsymbol{G}_q^H, \tag{5.21}$$

where

$$\mathsf{B}_{pq} = \boldsymbol{E}_p \mathsf{B} \boldsymbol{E}_q, \tag{5.22}$$

5.3.3 Implementing the RIME with MeqTrees

MeqTrees [148] is a sophisticated software package which allows users to implement the RIME to build models for simulations and also provides some calibration tools. The name MeqTrees derives from the underlying structure of the package: models are derived from a *measurement equation* and these are specified using *trees*, which is a flexible way to define mathematical expressions in code.

In this context, a tree is a graph whose *nodes* are connected in a parent-child hierarchy. Figure 5.3 shows an example tree for the expression $\sin a + \cos b$. The a and b nodes are called *leaf* nodes because they have no children, while the $+$ node is referred to as the *root* node as it has no parent. The tree is evaluated by starting at the leaves and propagating the values through their parents, performing all the relevant mathematical expressions, until the root is reached. A collection of trees, which may be interlinked, is called a *forest*. A node can be any kind of function (a constant, a polynomial, a FITS[3] image of the sky, a matrix, for example) making trees a highly flexible and object oriented way of implementing the RIME, or any kind of calculation, in MeqTrees. Figure 5.4 illustrates a sub-tree implementing the general RIME for N point sources.

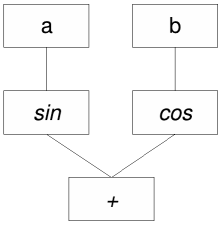


FIGURE 5.3: An example expression tree[148].

MeqTrees consists of several components, ranging from a high level user interface to a low level computational back-end. The *meqbrowser* is the GUI (graphical user interface) front-end which allows the user to easily generate new simulations and provides visualisations. The *meqserver* is the back-end which does the heavy computations, written largely in C++. Users can write scripts in the Python-based scripting language called *Tree Definition Language (TDL)* to control simulations and can implement some of the already supplied, versatile *frameworks* which

[3]Flexible Image Transport System, http://fits.gsfc.nasa.gov/fits_home.html

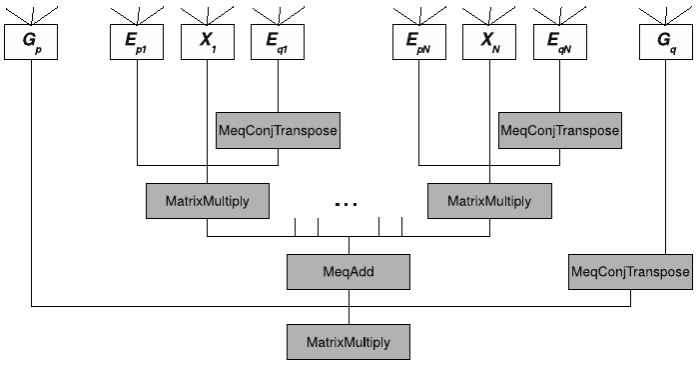


FIGURE 5.4: A tree implementing the RIME given in Eq.(5.20) [148].

implement many different measurement equations, simulations and calibration tools. There are also several ancillary tools for record keeping, image making etc.

To implement a model in MeqTrees, a user has to essentially provide it with a sky model and a set of Jones matrices, which MeqTrees then uses to evaluate the RIME and output the visbilities. In practice, there are several sky models already provided within MeqTrees which one can use and edit within a TDL file and similarly, many Jones matrices exist preprogrammed for common direction dependent and direction independent effects for telescopes including the WSRT and LOFAR[4]. Using the simple Tree Definition Language, the user can combine these existing models and Jones matrices, or using Python, can create their own to simulate any sky, telescope and interfering effects.

5.4 The BIRO Algorithm

In order to implement any kind of Bayesian analysis, a model for the data is required. In the case of radio interferometry, a rather sophisticated model is generally required. Fortunately, with the development of the radio interferometry measurement equation (see Section 5.3), this has become a relatively straightforward task. MeqTrees [148] (Section 5.3.3) is general software developed for modeling any kind of radio field with any kind of radio interferometer. MeqTrees takes from the user a sky model (such as the number and distribution of sources, their fluxes, shapes etc.) as well as instrumental details (such as the telescope configuration, beam pattern, pointing errors, noise, atmospheric effects etc.) and uses the measurement equation to produce realistic simulated visibilities that such a telescope would observe. MeqTrees has been useful for predicting the capabilities for future experiments and for understanding the intricacies of

[4]Low-Frequency Array http://www.lofar.org/

current telescopes. In this chapter, we go a step further and use MeqTrees as a modeling step in our Bayesian analysis.

MeqTrees is like any model, it takes in parameters and produces output which can be directly compared with data. Thus we can use it in an MCMC analysis (see Section 3.3.1 for details) to determine the posterior for the parameters. For this work, we first simulate a dataset with MeqTrees (see Section 5.6). Here we have a model for how the sky should look for this simulated dataset but in practice, it may not be so obvious how to select the sky model for real data. The Bayesian evidence (see Section 3.4.1) gives us the ability to select between models and indicate whether a model is performing badly and there are several ways one can imagine getting an initial idea of what the field may look like (starting with CLEAN or RESOLVE for example). Still it is worth noting that most Bayesian techniques will yield poor results if a poor model is selected.

With MeqTrees set up using the correct sky and telescope model, we are able to repeatedly make calls to it with different sets of parameters. Thus for every step of the MCMC chain, MeqTrees is called with a new set of parameters and produces a new set of visibilities to be directly compared with the simulated data, to compute the likelihood. It is this powerful combination of two separate pieces of software, MeqTrees and MCMC, which allows BIRO to work in a wide variety of possible cases. Figure 5.5 shows a simple schematic overview of the BIRO approach.

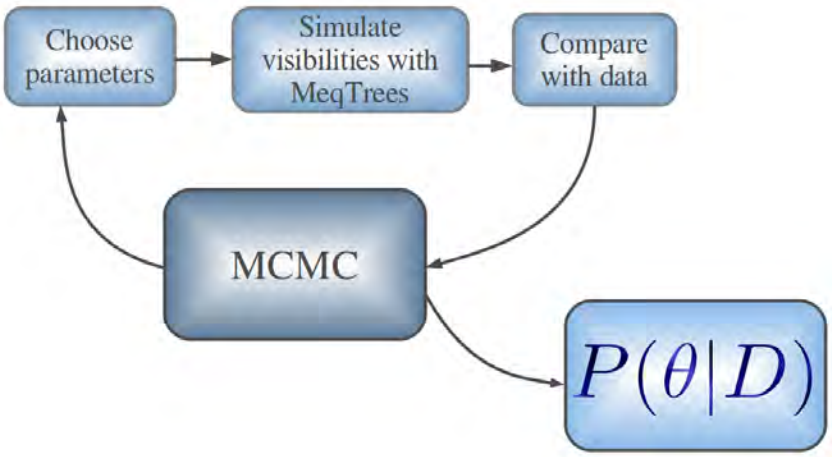


FIGURE 5.5: An overview of Bayesian Inference for Radio Observations. For any given dataset the MCMC algorithm iterates choosing a set of parameters, simulating visibilities with those parameters using MeqTrees and comparing the results with the data. Good fits are accepted and poor fits are discarded. In this way, the MCMC builds up a good approximation to the full posterior, $P(\theta|D)$, that is, the probability of the parameters given the data, which contains the estimates of the parameters and their uncertainties and correlations.

We implemented BIRO for three different simulated datasets, which we had complete control over, to test the technique. In each case, we used the Tree Definition Language (TDL) configuration files to define the model (for example, choosing a known beam model or providing pointing errors simply as a list of numbers), which allowed a layer of abstraction between the user and the underlying workings of the RIME. We used the "makems" package (available as part of MeqTrees) to first describe and create the measurement set file before simulating the data. We chose to simulate WSRT[5] data, due to it being well supported already within MeqTrees.

After simulating a dataset, we used the same TDL to create a model whose parameters we wish to determine. Thus we use the same software to generate the dataset and to produce the model, meaning any problems we encounter should be indicative of issues with the technique rather than the data. We used the same Python[6] MCMC code developed for the work done in Chapter 4 to repeatedly call MeqTrees, varying the parameters of the model to determine the full posterior.

For each dataset, we initially ran 5 chains of at least 300 000 steps each recalculating the covariance matrix every 25 000 steps in order to get a good estimate. We then obtained one estimate of the covariance matrix from adding together these multiple chains and restarted them with a fixed covariance matrix in order to obtain the final results. It typically took at least 300 000 steps per chain to ensure convergence. Convergence was tested in all cases with the Gelman-Rubin statistic (see Section 3.3.1) and by visually inspecting the density plots of the chains to ensure they were indistinguishable, meaning they had all converged to the same distribution. Figure 5.6 shows an example density plot from one of the datasets.

Worth some discussion is the starting point of these chains. Usually, MCMC chains should start at any random point in the parameter space. However, in this case because of the high dimensionality of the problem (up to 153 parameters), we found this was unpractical as the chain would take too long to get to a good starting point where most of the probability is located. We found that the Broyden-Fletcher-Goldfarb-Shanno optimisation[7] algorithm, as implemented in Python, does a good job in getting a close initial guess to the true optimum to use as a starting point for the MCMC. The algorithm is not very fast, requiring several thousand likelihood function calls, but was far more successful than stochastic algorithms tried such as simulated annealing and always got reasonably close to the optimum from a random position in a very

[5] Westerbork Synthesis Radio Telescope, https://www.astron.nl/radio-observatory/astronomers/wsrt-astronomers

[6] www.python.org

[7] http://en.wikipedia.org/wiki/Broyden-Fletcher-Goldfarb-Shanno_algorithm

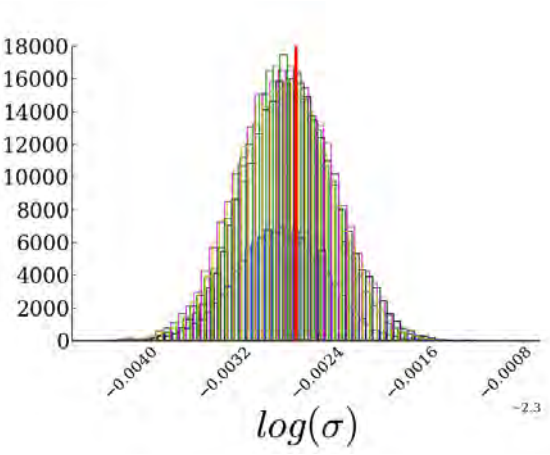


FIGURE 5.6: An example density plot for one of the parameters from the grid dataset (see 5.6.1). Each of the ten colours represents a different chain.

large parameter space. We note that such optimisation algorithms can be used to get a good approximation to before starting an MCMC chain.

5.4.1 Bayesian Factor Graphs

Before discussing the simulated datasets and the results from BIRO, we first introduce Bayesian factor graphs, useful tools for visualising Bayesian models, which we will use in describing the data and models applied in the next section.

We utilise the directed factor graph notation, developed in [149], to visualise how the parameters in our models depend on one another. Table 5.1 defines the graphical primitives of a factor graph. Figure 5.7 demonstrates the use of the factor graph notation in a simple example.

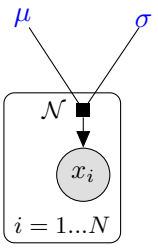


FIGURE 5.7: A simple example factor graph. In this model, the data is represented by a vector x_i, which we suspect is normally distributed. This is modeled by a normal distribution (represented by the factor labeled $\mathcal{N}$) which is governed by the parameters μ and σ. These constants would be the parameters we would want to estimate with an MCMC analysis.

5.5 Applying BIRO

In applying BIRO for the first time, we wanted to focus on the relationships between scientific and instrumental parameters and perform a joint analysis. Here we present the general model used in our simulated datasets.

Our simulated scientific parameters consisted of fluxes of sources and, in two of the simulated datasets, the shape parameters of extended sources. The calibration errors we focused on were pointing errors, the beam width of the primary beam and the noise on the visibilities. Below we derive a general measurement equation for this class of models. For each specific model and simulated dataset in the next section, these are illustrated with Bayesian factor graphs.

TABLE 5.1: Factor graph node types (adapted from [149]).

Type	Description	Output
Latent variable	An unobserved, undetermined variable	x
Observed variable	The data	y
Deterministic	A variable completely determined by other variables	$\boldsymbol{Z}$
Constant	A "true" variable. We usually want to determine these	a
Factor	A distribution to be drawn from	$\mathcal{N}$ ■
Plate	Repeats variables or factors	x_m $m \in \mathcal{M}$

5.5.1 The Brightness Matrix for Extended Sources

We only consider unpolarised sources for this work. From Eq.(5.10), the brightness matrix of an unpolarised point source is:

$$\mathsf{B}^{\text{POINT}} = \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix} \tag{5.23}$$

One can apply the RIME for extended emission as well. We need an expression for the distribution of the emission as a function of direction which we can substitute for B in the full-sky RIME of Eq.(5.21).

For example, the brightness matrix for an unpolarised, Gaussian extended source is given by:

$$\mathsf{B}^{\text{GAUSS}} = f(l,m) \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix}, \tag{5.24}$$

where I is the flux of the source.

The function, $f(l,m)$ is a two-dimensional Gaussian function which, in general, is written as:

$$f(x,y) = A \exp\big[-(a(x-x_0)^2 + 2b(x-x_0)(y-y_0) + c(y-y_0)^2)\big], \tag{5.25}$$

where (x_0, y_0) is the centre of the Gaussian and the coefficients a,b and c are given by:

$$a = \frac{\cos^2(\theta)}{2\sigma_x^2} + \frac{\sin^2(\theta)}{2\sigma_y^2} \tag{5.26}$$

$$b = \frac{\sin(2\theta)}{4\sigma_x^2} - \frac{\sin(2\theta)}{4\sigma_y^2} \tag{5.27}$$

$$c = \frac{\sin^2(\theta)}{2\sigma_x^2} + \frac{\cos^2(\theta)}{2\sigma_y^2}. \tag{5.28}$$

Here, θ is the angle of the Gaussian from the x-axis in the anti-clockwise direction and σ_x and σ_y are the widths in the x and y direction (at least, they would be in the x and y direction if θ were set to zero).

Further, a Gaussian extended source is parameterised in MeqTrees as the following:

$$l_\perp = e_{\text{maj}} \sin(\alpha) \tag{5.29}$$

$$m_\perp = e_{\text{maj}} \cos(\alpha) \tag{5.30}$$

$$r = e_{\text{min}}/e_{\text{maj}} \tag{5.31}$$

where e_{maj} and e_{min} are the major and minor axes of the Gaussian source and α is the position angle. Linking this to Eq.(5.25) and (5.26): α corresponds to θ and e_{maj} is the maximum of (σ_x,σ_y) while e_{min} is the minimum of the two. Figure 5.8 illustrates this parameterisation of a Gaussian extended source.

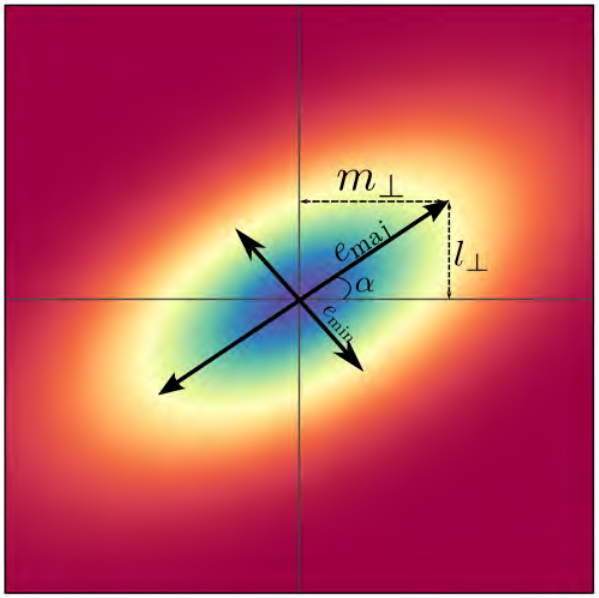


FIGURE 5.8: The parameterisation of a Gaussian extended source in MeqTrees.

5.5.2 The Primary BEAM

The effects of the primary beam can be included in the Jones' matrices of the RIME. For example, the standard model often used for the primary beam of WSRT[8] is given by[9]:

$$\cos^3(cnr), \tag{5.32}$$

where n is the observing frequency (in GHz), r is the distance from the pointing centre in degrees and c is the beam factor (in 1/GHz). This enters into the RIME as [150]:

$$E^{\text{BEAM}}(l,m) = \cos^3(c\nu\sqrt{l^2+m^2}) \tag{5.33}$$

where for WSRT, l is the right ascension direction and m is the declination direction.

[8] Westerbork Synthesis Radio Telescope, https://www.astron.nl/radio-observatory/astronomers/wsrt-astronomers

[9] https://www.astron.nl/radio-observatory/astronomers/wsrt-guide-observations/5-technical-information/5-technical-informatio, note that in this link the power beam is considered which is the "voltage" beam squared.

FIGURE 5.9: A simple illustration of how a telescope pointing error could affect the measurement of flux in a point source. *Left panel:* Antenna without a pointing error. *Right panel:* Antenna with a pointing error, in which the flux of the source is decreased due to the mispointing. Photo credit: Nadeem Oozeer.

5.5.3 Pointing Errors

Pointing errors occur when an individual antenna mispoints slightly at its target, due to effects such as gravitational tugging at the structure as it moves and wind effects. Because these errors are small and the beam of each antenna is usually fairly large, we'd expect the largest effect of pointing errors to be on the fluxes of the sources, as the source is seen through the side of the beam instead of through the centre, as illustrated in Figure 5.9. Of course, the situation may become more complex when multiple sources are considered [151].

Pointing errors can be easily incorporated as per antenna, direction dependent Jones matrices [150]:

$$E_p^{\mathrm{PE}}(l,m) = E(l+\delta l_p, m+\delta l_p), \tag{5.34}$$

where δl_p and δm_p are the pointing errors for antenna p in the right ascension and declination direction respectively.

5.5.4 The Measurement Equation

Combining these Jones' and brightness matrices into the full-sky RIME of Eq.(5.21), our general model is:

$$\mathsf{V}_{pq} = \boldsymbol{G}_p \left(\sum_s \left(\int \int_{lm} \mathsf{B}_{pq} \mathrm{e}^{-2\pi i(u_{pq}l + v_{pq}m)} \mathrm{d}l\mathrm{d}m \right) \right) \boldsymbol{G}_q^H, \tag{5.35}$$

where

$$\mathsf{B}_{pq} = E^{\mathrm{BEAM}}(l,m) \left(\boldsymbol{E}^{\mathrm{PE}}{}_p \mathsf{B}_s \boldsymbol{E}^{\mathrm{PE}}{}_q \right) E^{\mathrm{BEAM}}(l,m), \tag{5.36}$$

where B_s is the brightness matrix for source s, either a Gaussian or point source.

5.5.5 The Likelihood

The likelihood, $P(D|\theta)$, which defines how well a given set of parameters, θ, fits the data, D, is given by:

$$\mathcal{L} = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-\chi^2}{2} \right) \tag{5.37}$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(\widetilde{\mathsf{V}}_{pq} - \mathsf{V}_{pq})^2}{2\sigma^2} \right), \tag{5.38}$$

where V_{pq} are the observed visibilities, $\widetilde{\mathsf{V}}_{pq}$ are the theoretical visibilities produced by MeqTrees and σ is the noise on the visibilities. Since the noise in our simulations is Gaussian, it follows that we should use a Gaussian likelihood. It is generally assumed that the noise is the same for all visibilities but it may be possible to have multiple values of σ if, for example, each antenna has a different system temperature and thus a different sensitivity. In this case, the noise for a given baseline is the geometric mean of the noise of each antenna in the baseline and the likelihood would need to be written as the product of likelihoods for each individual baseline.

5.6 The Datasets

To prove BIRO works and demonstrate its superior properties over a standard CLEAN and source extraction approach, we tested it on three, fully understood simulated datasets of increasing complexity. The datasets were simulated using MeqTrees. They are described as:

- **Grid dataset -** This dataset consists of 9 point sources on a 3x3 grid with constant pointing errors in the antennae. For this model, there are 39 parameters we fit for: 9 scientific, 30 nuisance parameters.

- **QMC dataset -** QMC stands for Quality Monitoring Committee. This more complex dataset is based on a real WSRT field sometimes used for calibration due to its convenient arrangement of point sources. Our simulated dataset includes 16 point sources and 1 Gaussian extended source, testing to see if our algorithm can simultaneously recover unknown point source and extended source parameters. We again added constant pointing errors to the antennae. For this model, there are 50 parameters we fit for: 20 scientific, 30 nuisance parameters.

- **QMC shape dataset -** We took the previous dataset and allowed for the possibility of all the sources being extended to see whether we could recover the unknown shape parameters and *distinguish* between point and extended sources. For this model, there are 98 parameters we fit for: 68 scientific, 30 nuisance parameters.

5.6.1 Grid Mock Data

The field has nine point sources of known positions and we wish to determine the fluxes of the sources (these are our science parameters) and the uncertainty on those fluxes.

In addition to the science parameters, we added some realistic instrumental parameters to deal with. We include pointing errors, as discussed in Section 5.5.3. Because these errors are small and the beam of each antenna is usually fairly large, we'd expect the largest effect of pointing errors to be on the fluxes of the sources, as the source is seen through the side of the beam instead of through the centre. Thus we can immediately suspect there will be a correlation between our scientific and instrumental parameters. However, pointing errors are often ignored or simply estimated as best as the astronomer can and then fixed to this initial value instead of estimated directly from the data. So we injected 28 pointing errors, two for each of the fourteen WSRT antennae (one for each independent direction), to our simulated dataset.

We also fit for c, the beam width of the WSRT primary beam, as discussed in Section 5.5.2 Of course this model of the beam is quite simple and one could provide a more complex model and easily fit those parameters with this technique as well.

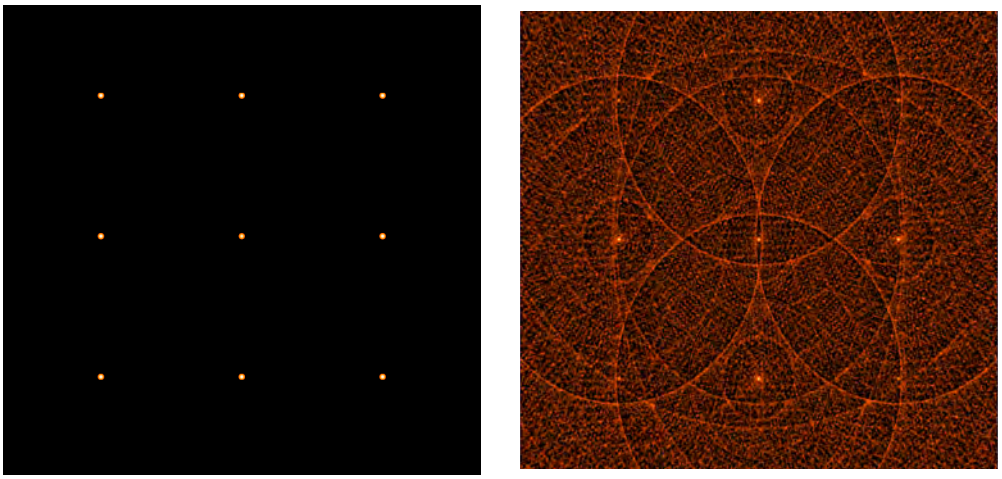


FIGURE 5.10: *Left panel:* The "grid" simulated sky model. *Right panel:* How this field appears when convolved with the telescope beam, with all parameters set to their true values. The rings (sidelobes) of the dirty beam are clearly visibile from the telescope's point source response, caused by an incomplete sampling of the uv-plane (see Figure 2.6).

Lastly, as the noise on the visibilities will generally be unknown but will likely be Gaussian, we included one final parameter for the standard deviation of the noise on the visibilities (which is varied in log-space to allow it to be as arbitrarily close to zero as the data requires). This brings our total number of parameters for the model which describes our mock dataset to 39. Figure 5.10 shows the input model and the "dirty" image, as the true model appears when convolved with the telescope's beam.

Figure 5.11 illustrates the model used to generate and analyse this dataset with a Bayesian factor graph (see Section 5.4.1). The measurement equation used is identical to that discussed in Section 5.5 in Eq.(5.35) and (5.36), where the pointing errors δl and δm are constant offsets and the brightness matrices are all point sources:

$$\mathsf{V}_{pq} = \boldsymbol{G}_p \left(\sum_s \left(\int\int_{lm} \mathsf{B}_{pq} \mathrm{e}^{-2\pi i(u_{pq}l + v_{pq}m)} \mathrm{d}l\mathrm{d}m \right) \right) \boldsymbol{G}_q^H, \tag{5.39}$$

$$\text{where } \mathsf{B}_{pq} = E^{\mathrm{BEAM}}(l,m) \left(\boldsymbol{E}^{\mathrm{PE}}{}_p \mathsf{B}_s^{\mathrm{POINT}} \boldsymbol{E}^{\mathrm{PE}}{}_q \right) E^{\mathrm{BEAM}}(l,m). \tag{5.40}$$

Constant priors between -200 and 200 arcseconds were applied to the pointing errors and a positive prior was applied to the beam width.

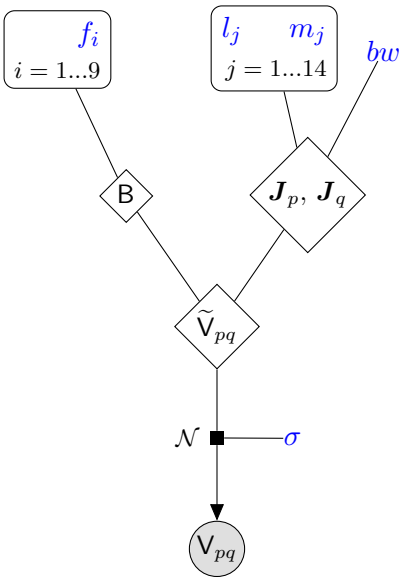


FIGURE 5.11: Factor graph of the "grid" model with constant pointing errors. All parameters estimated with MCMC are the constants, without any circles around them, coloured blue. The V_{pq} are the observed visibilities, drawn from a normal distribution of mean $\widetilde{\mathsf{V}}_{pq}$ (the unobserved, true visbilities) and standard deviation σ. These "true" visibilities are governed by the RIME, which is here simplified graphically to two components, the brightness matrix, B, and the Jones' matrices of the antennae, $\boldsymbol{J}_p$, $\boldsymbol{J}_q$. The fluxes of the nine sources are represented by f_i, which form components of B, and the pointing errors, l_j and m_j enter the Jones matrices, along with the beam width, bw.

5.6.1.1 Results

Figure 5.12 shows the 1d marginalised posterior for two of the parameters for the "grid" dataset, demonstrating that we are able to accurately determine the true value of each parameter to within the estimated uncertainty. All the histograms for this dataset are in Appendix A. To our knowledge, accurate estimates and uncertainties on pointing errors have never been established before now. We are able to recover the best fit and uncertainty on all the parameters, both scientific and calibration, simultaneously. We are able to achieve an unbiased estimate for the fluxes as well as place error bars on them that we can truly believe, error bars which include rather than ignore all known sources of instrumental interference. The beam width is well estimated, showing several different types of calibration errors can be recovered. What is also impressive is BIRO's ability to accurately recover the noise on the visibilities, with an estimate of the uncertainty.

It is essential to compare any new technique with existing techniques. We did a blind comparison by giving our dataset, without any other information, to a colleague experienced in standard radio astronomy techniques, and asked him to determine the fluxes of the sources. He utilised

a program called PyBDSM[10] to first apply CLEAN to the data and then extract the sources. The PyBDSM analysis found 31 sources, determining that almost all of them were extended. We found the 9 sources whose positions best match the positions of the true sources and then compared their fluxes (see Table A.1 in Appendix A). While the image produced by CLEAN certainly looks reasonable, the difference between the clean image and the true model in Figure 5.13 shows that although the positions of the sources are correct, their fluxes are clearly wrong. Figure 5.14 shows a numerical comparison in the relative errors of the fluxes between BIRO and CLEAN+SE (that is, CLEAN followed by source extraction). The true values for the fluxes are consistent with the BIRO error bars (the zero line on this relative error plot), while the CLEAN+SE results are dramatically biased, in some instances by over 100-σ!

This most basic application of CLEAN+SE probably produces such biased results because it cannot gain any knowledge about pointing errors by itself from the data and thus cannot correct for them. In general, one could get better flux estimates by iterating between calibration and

[10] http://dl.dropboxusercontent.com/u/1948170/html/index.html

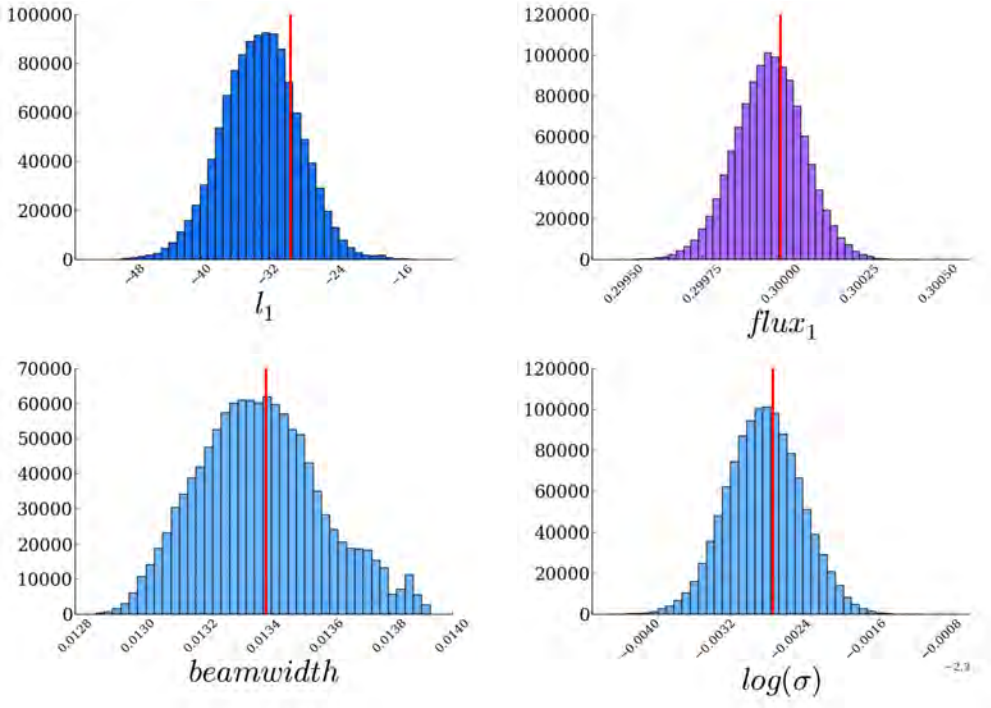


FIGURE 5.12: Selected 1d marginalised posteriors from the MCMC analysis of the "grid" dataset, demonstrating BIRO's ability to recover the joint posterior for scientific (i.e. the flux) and instrumental (i.e. pointing error, beam width and noise) parameters with fully marginalised errors. The true value is given by the red vertical line. Parameters are (reading from left to right, top to bottom) l_1: the pointing error in the right ascension direction for the first antenna, $flux_1$: the flux of the central source, $beamwidth$: parameter c in Eq.(5.32) and $log(\sigma)$: the log of the standard deviation of the visibility noise.

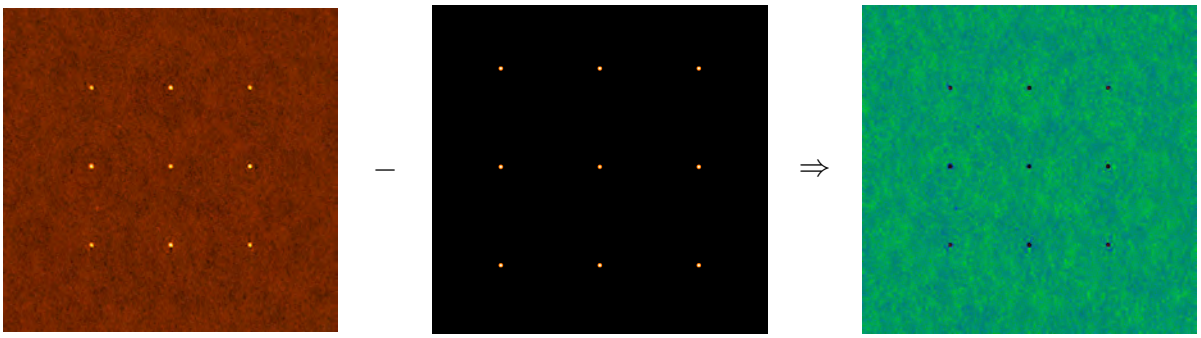


FIGURE 5.13: CLEAN applied to the "grid" dataset. *Left panel:* The image produced by applying CLEAN to the "grid" dataset. *Middle panel:* The true model. *Right panel:* The difference between the clean image and the true model. See Figure 5.14 for a numerical comparison.

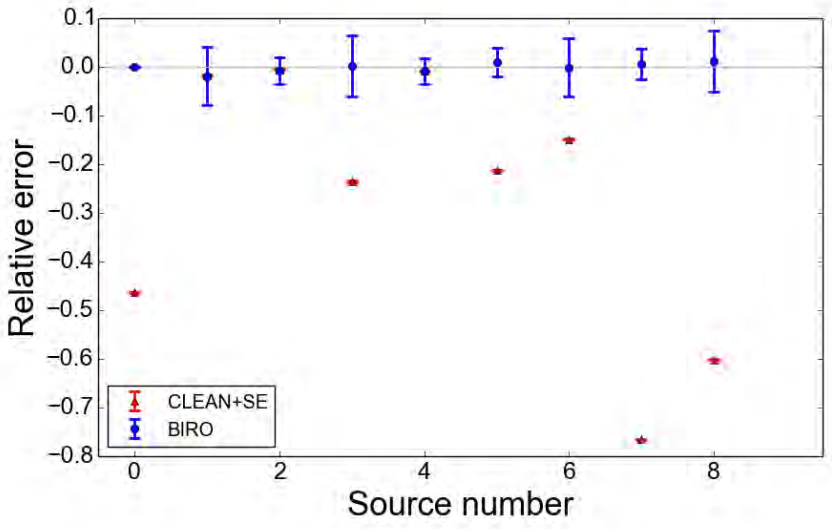


FIGURE 5.14: Comparison between BIRO and the results from CLEANing and then applying a source extraction algorithm to the data (CLEAN+SE). For each of the 9 sources in the dataset, we subtract the estimated flux from the true flux and divide the result by the true flux to obtain and plot the relative error (y-axis). We similarly scale the error bars for both BIRO and CLEAN+SE by dividing them by the true flux. While BIRO estimates the fluxes with a maximum error of 3%, CLEAN+SE is biased by up to 76%, due to the strong correlations between pointing error and flux.

source extraction to correct for the pointing errors. However, if these estimates are biased the resulting fluxes would still be biased and current calibration algorithms do not allow for more complex models of pointing errors such as time-varying pointing errors. The fluxes are all biased low, which is possibly due to the pointing errors (although may be due to other correlations as well) since, as described previously, if there is a pointing error, the source tends to lie in the edge instead of the middle of the primary beam. Pointing errors are not very well studied [151] but we can see from these results that they have a strong influence on scientific parameters returned by CLEAN+SE. In this case, it is at least clear that the error bars produced by PyBDSM are

underestimated and unreliable. This emphasises the usefulness of BIRO in its ability to estimate the pointing errors directly from the data itself.

Figure 5.15 shows a representation of the covariance matrix of the full posterior of the problem. This image is a overview of how all the parameters are correlated. We wrote code in R [152] to generate a normalised covariance matrix from the MCMC chains. The caption of the figure explains what the colours and shapes of the ellipses mean. The first thing to note about this plot is that, contrary to previous assumptions where pointing errors were fixed and then ignored, there are strong correlations between instrumental and scientific parameters. There are also complex interactions between the fluxes and between the pointing errors themselves. In fact, in this case the noise on the visibilities is the only parameter completely uncorrelated with anything else (as can be seen by the rows and columns of white circles).

These correlations arise due to the complexity of the deconvolution problem. The pointing errors are strongly correlated in each direction (although not between directions). Each pointing error affects every source, so in turn when one determines the pointing errors from the data, one finds correlations and degeneracies in the pointing errors. The fluxes are highly correlated due to the fact that there are gaps in the uv-coverage (the deconvolution problem) which places uncertainty on the flux distribution. The interaction between pointing errors and fluxes is obviously highly complex from looking at Figure 5.15. With multiple sources and multiple pointing errors, certain fluxes correlate positively, negatively or not at all with certain pointing errors. Fortunately, the power of BIRO means that one does not have to predict this covariance matrix from first principles (indeed it may well be impossible) but can estimate it directly from the data and learn about the relationships between parameters that way. Lastly, it should be noted that this covariance matrix is unique to WSRT. A completely different matrix would be generated for a different telescope, which would enable new analysis for that particular telescope.

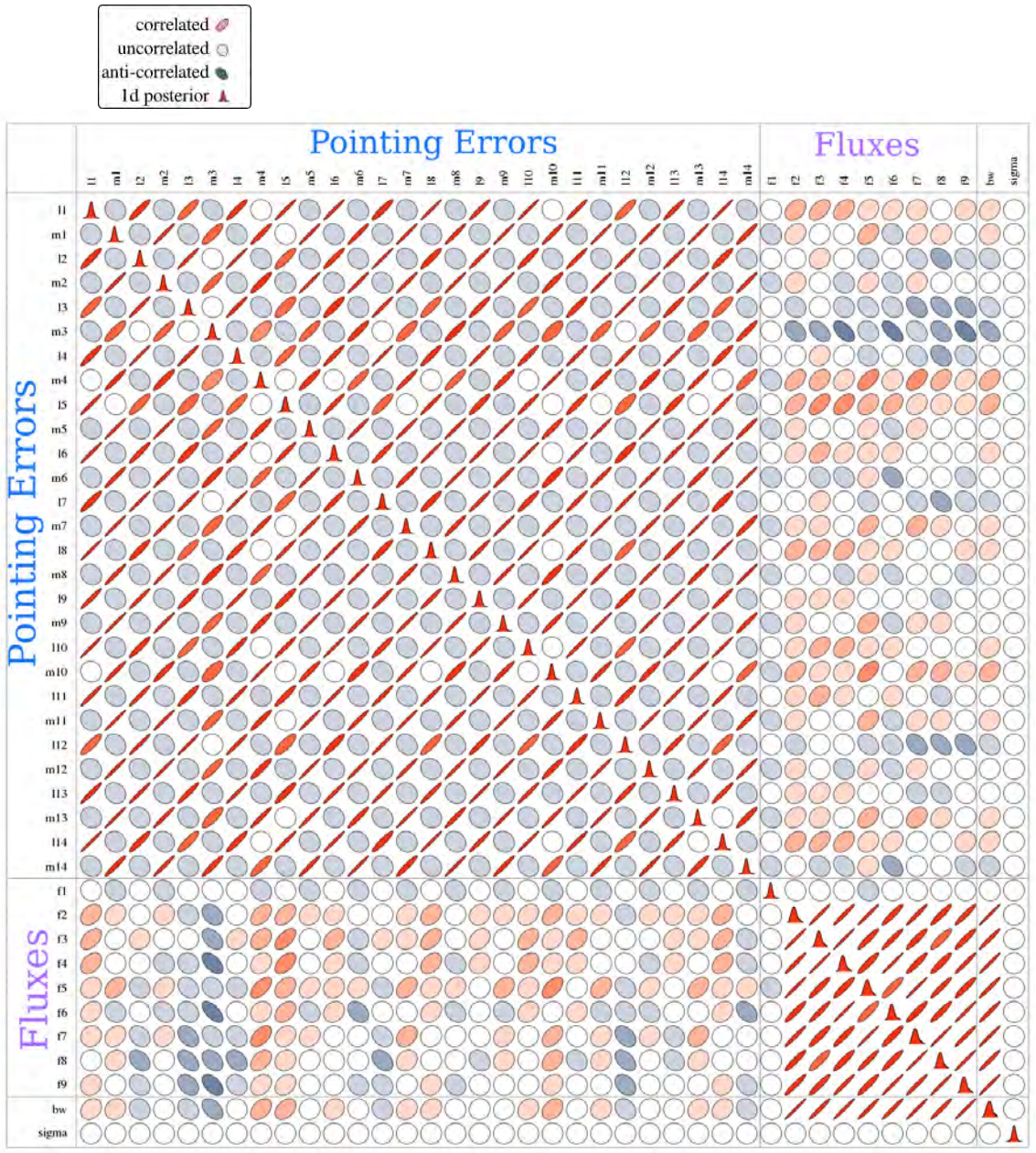


FIGURE 5.15: Covariance matrix between parameters for the 9-source simulated dataset. The parameters are listed on each axis (pointing errors for each antenna first, followed by the fluxes, then the beam width and the log of the noise) with the correlations between them represented by a coloured ellipse. Highly correlated parameters are red with a thin, angled ellipse, whereas anti-correlated parameters have a dark blue ellipse, angled to the left. The diagonal shows the 1d marginalised posterior for each parameter (the same as the above histograms).

5.6.2 QMC Mock Dataset

The QMC field[11] is more realistic than the previous dataset and the fluxes of the seventeen sources within it are again our scientific parameters we wish to estimate. In addition, we make one of the sources slightly extended, which we model as a Gaussian and then fit for these extended parameters.

We injected the same pointing errors as we did in the "grid" dataset into the "QMC" dataset and also applied the same beam model and fit for the beam width. We fit for the noise on the visibilities, bringing the final number of parameters for this dataset to 50. Figure 5.16 shows the input model and the "dirty" image, as the true model appears when convolved with the telescope's beam.

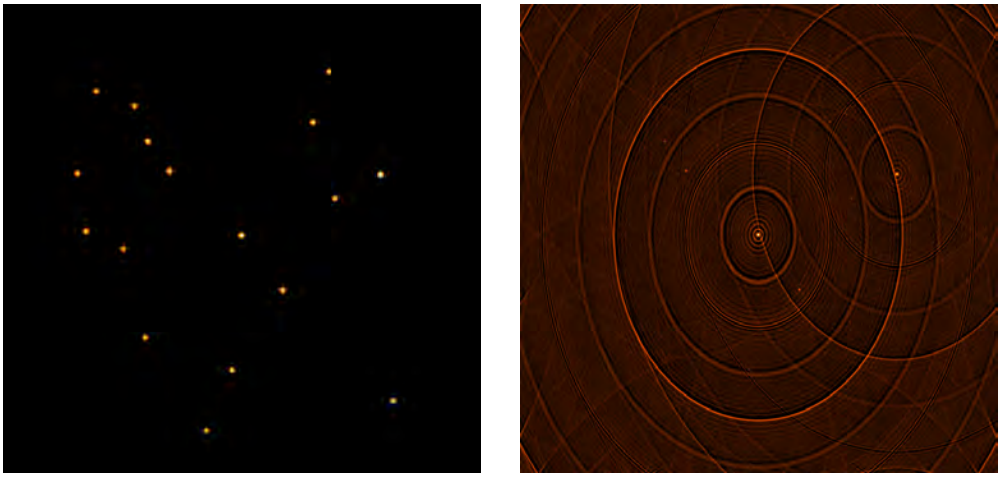


FIGURE 5.16: *Left panel:* The "QMC" simulated sky model. *Right panel:* How this field appears when convolved with the telescope beam, with all parameters set to their true values, where again the rings of the telescope's sidelobes are visible (see Figure 2.6).

Figure 5.17 illustrates the model used to generate and analyse this dataset with a Bayesian factor graph (see Section 5.4.1). The measurement equation used is similar to that discussed in Section 5.5 in Eq.(5.35) and (5.36), where the pointing errors δl and δm are constant offsets and the brightness matrices are all point sources, except for the first source, which is extended. This is illustrated in Eq.(5.42). Constant priors between -200 and 200 arcseconds were applied to the pointing errors and a positive prior was applied to the beam width. For this dataset, the three shape parameters for the extended source were also restricted with a positive prior.

[11] QMC stands for Quality Monitoring Committee and this is a sometimes used simulated dataset internally amongst the WSRT calibration team.

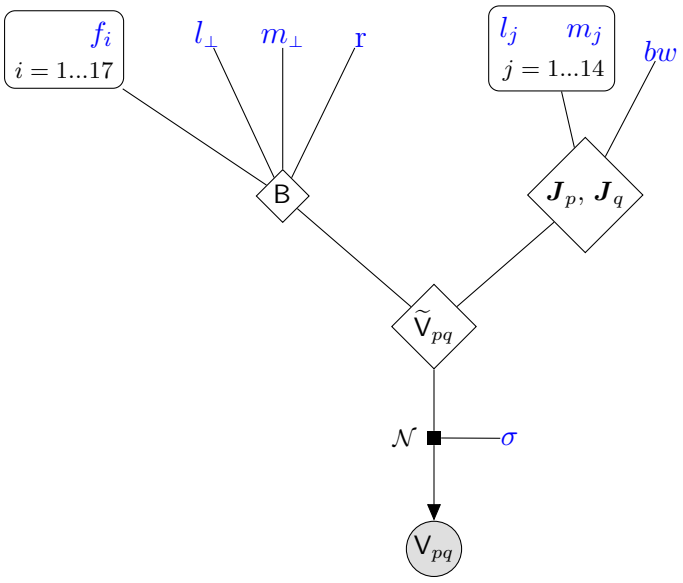


FIGURE 5.17: Factor graph of the "QMC" model with constant pointing errors, similar to Figure 5.11. All parameters estimated with MCMC are the constants, without any circles around them, coloured blue. The V_{pq} are the observed visibilities, drawn from a normal distribution of mean $\widetilde{\mathsf{V}}_{pq}$ (the unobserved, true visbilities) and standard deviation σ. These "true" visibilities are governed by the RIME, which is here simplified graphically to two components, the brightness matrix, B, and the Jones' matrices of the antennae, $\boldsymbol{J}_p$, $\boldsymbol{J}_q$. The fluxes of the seventeen sources are represented by f_i, which form components of B, and $l_\perp$, $m_\perp$ and r are the shape parameters of the first source, the only extended source. The pointing errors, l_j and m_j enter the Jones matrices, along with the beam width, bw.

$$\mathsf{V}_{pq} = \boldsymbol{G}_p \left(\sum_s \left(\int\int_{lm} \mathsf{B}_{pq} \mathrm{e}^{-2\pi i(u_{pq}l + v_{pq}m)} \mathrm{d}l\mathrm{d}m \right) \right) \boldsymbol{G}_q^H, \tag{5.41}$$

$$\text{where } \mathsf{B}_{pq} = E^{\mathrm{BEAM}}(l,m) \left(\boldsymbol{E}^{\mathrm{PE}}{}_p \mathsf{B}_s \boldsymbol{E}^{\mathrm{PE}}{}_q \right) E^{\mathrm{BEAM}}(l,m). \tag{5.42}$$

The brightness matrix for the first source will be a Gaussian (see Section 5.5.1), $\mathsf{B}^{\mathrm{GAUSS}}$, and a point source, $\mathsf{B}^{\mathrm{POINT}}$, for all the other sources.

5.6.2.1 Results

Figure 5.18 shows the 1d marginalised posterior for three of the parameters for the "QMC" dataset, demonstrating that we are able to accurately determine the true value of each parameter to within the estimated uncertainty. All the histograms for this dataset are in Appendix A. There are mild biases within $1-\sigma$ or $2-\sigma$ of the true values, which one always expects due to random noise in the dataset. These biases do not appear to be systematic between dataset realisations. We are able to not only extract the pointing errors and fluxes of the point sources,

but also the shape parameters of the extended source. While CLEAN (in its basic form) has to assume the map is made up of point sources, we need not make such an assumption, but can in theory model a field of any combination of point and extended sources of any shape that can be included in MeqTrees.

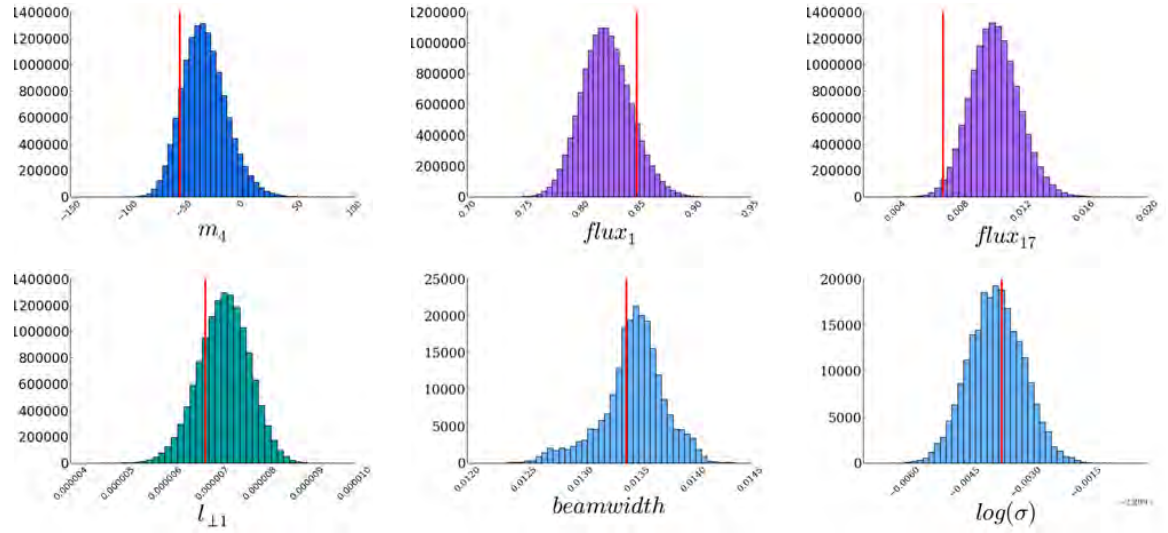


FIGURE 5.18: Selected 1d marginalised posteriors from the MCMC analysis of the "QMC" dataset. The true value is given by the red vertical line. These plots illustrate how we simultaneously estimate the flux of the point sources, an extended source, the shape of the extended source and various instrumental effects. Parameters are (reading from left to right, top to bottom) m_4: the pointing error in the declination direction for the second antenna, $flux_1$: the flux of the brightest source, $flux_{17}$: the flux of the faintest source, $l_\perp$: the projection of the major axis of the extended source on the RA-direction, *beamwidth*: parameter c in Eq.(5.32) and $log(\sigma)$: the log of the standard deviation of the visibility noise.

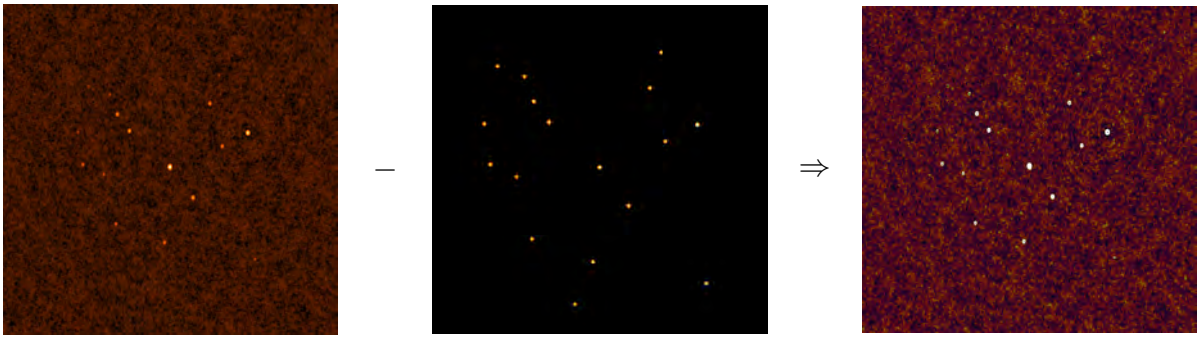


FIGURE 5.19: CLEAN applied to the "QMC" dataset. *Left panel:* The CLEANed image. *Middle panel:* The true model. *Right panel:* The difference between the clean image and the true model. See Figure 5.20 for a numerical comparison.

Once again we did a blind comparison by giving our dataset, without any other information, to a colleague experienced in standard radio astronomy techniques, and asked him to determine the fluxes of the sources. The PyBDSM analysis found 20 sources, determining that many of the point sources were extended. We found the 17 sources whose positions best match the positions of the true sources and then compared their fluxes (see Table A.2 in Appendix A).

Again, whilst the image produced by CLEAN seems reasonable, taking the difference between it and the model, as in Figure 5.19 reveals the image is not representative of the true sky. Interestingly, CLEAN+SE fits Gaussian shapes to the point sources as is clear in the right panel of that figure.

Figure 5.20 shows a comparison in the relative errors of the fluxes between BIRO and CLEAN+SE (that is, CLEAN followed by source extraction). BIRO is able to recover each of the 17 fluxes to within 1-σ of their true value (the zero line on this relative error plot), while the CLEAN+SE results are biased for the brighter sources. The results for this dataset are not as dramatic as there are fewer bright sources which are most strongly affected by pointing errors. Even so, the estimate of the flux of the brightest source by CLEAN+SE is wrong by up to 20%, but more important than this number is the fact that it is biased by over $100 - \sigma$. The error bars placed on the fluxes by PyBDSM do not include such sources of instrumental error and as such should be treated with caution. While BIRO is able to estimate the nuisance parameters (i.e. the pointing errors) directly from the data, the naive analysis with CLEAN+SE assumes they are zero and thus suffers these biases, and cannot include this assumption in its uncertainty estimate. The Bayesian approach, while still susceptible to human error, allows us to include all known sources of error and then marginalise naturally over these effects. It should also be

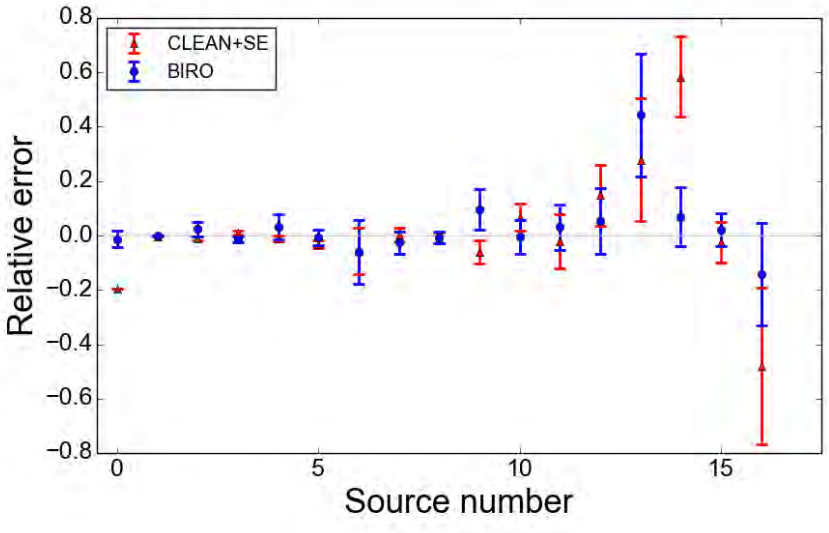


FIGURE 5.20: Comparison between BIRO and the results from CLEANing and then applying a source extraction algorithm to the data (CLEAN+SE). For each of the 17 sources in the dataset, we subtract the estimated flux from the true flux and divide the result by the true flux to obtain and plot the relative error (y-axis). We similarly scale the error bars for both BIRO and CLEAN+SE by dividing them by the true flux. While BIRO is able to estimate the true fluxes to within its estimated uncertainties, CLEAN+SE produces biased ($> 3\sigma$)results for several sources.

noted that while BIRO was able to recover the shape parameters of the one extended source, CLEAN+SE classified it as a point source.

Figure 5.21 shows the covariance matrix of the full posterior of the problem, using the same code used to produce Figure 5.15. The caption of the figure explains what the colours and shapes of the ellipses mean. Again there are strong correlations between instrumental and scientific parameters.

There are strong correlations between the fluxes and between the pointing errors. The interaction between pointing errors and fluxes is obviously highly complex from looking at Figure 5.21. Interestingly, there seems to be little interaction between the shape parameters and the other parameters. The beam width is highly correlated with almost all the other parameters, implying that modeling and understanding the beam is of crucial importance. The pointing errors are again highly correlated for this dataset since they interfere with every source. The fluxes are highly correlated due to the nature of the deconvolution problem and uncertainty due to missing information from gaps in the uv-plane. The noise of the visibilities is independent of all other parameters.

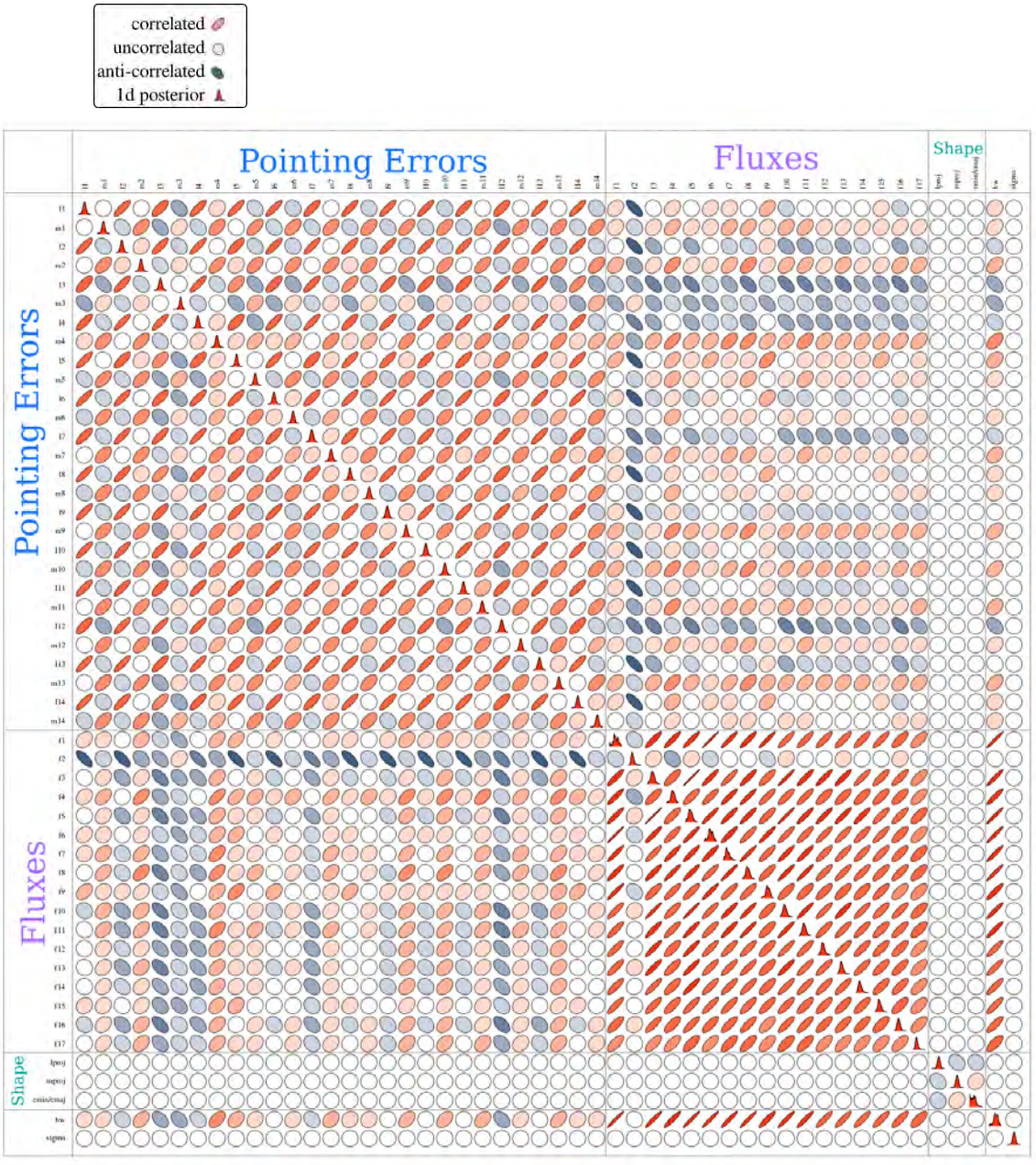


FIGURE 5.21: Covariance matrix between parameters for the "QMC" simulated dataset. The parameters are listed on each axis (pointing errors for each antenna first, followed by the fluxes, then the beam width and the log of the noise) with the correlations between them represented by a coloured ellipse. Highly correlated parameters are red with a thin, angled ellipse, whereas anti-correlated parameters have a dark blue ellipse, angled to the left. The diagonal shows the 1d marginalised posterior for each parameter (the same as the above histograms).

5.6.3 QMC Mock Dataset with Varying Shape Parameters

Here we use the same model as in the previous analysis but add more parameters. We realise that it is unrealistic to know *a priori* whether or not a given source is a point source or is in fact slightly extended. So we allowed the shape parameters of every source to vary. For a point source, the estimates of the shape parameters should all be zero. That means, all 17 sources will have $l_\perp$, $m_\perp$ and r parameters associated with them but we would expect that for the 16 point sources, these parameters would be zero. With the same 28 pointing errors, 17 fluxes, the beam width, noise and now 3×17 shape parameters, this model has 98 parameters to fit. The input model and dirty image are the same as Figure 5.16 and Figure 5.22 shows the factor graph for this model.

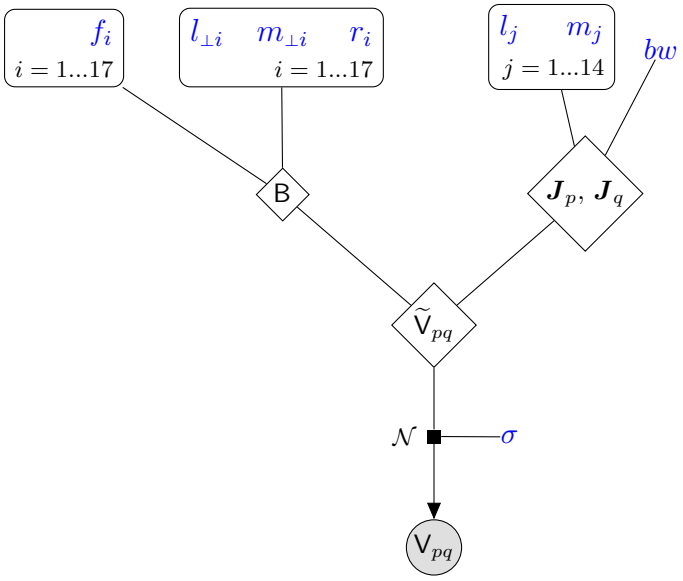


FIGURE 5.22: Factor graph of the "QMC shape" model with constant pointing errors, similar to Figure 5.17. All parameters we estimate with MCMC are the constants, without any circles around them, coloured blue. The V_{pq} are the observed visibilities, which are drawn from a normal distribution of mean $\tilde{\mathsf{V}}_{pq}$ (the unobserved, true visbilities) and standard deviation σ, which is one of the parameters we estimate with MCMC. These "true" visibilities are governed by the RIME, which is here simplified graphically to two components, the brightness matrix, B, and the Jones' matrices of the antennae, $\boldsymbol{J}_p$, $\boldsymbol{J}_q$. The fluxes of the seventeen sources are represented by f_i, which form components of B, and $l_{\perp i}$, $m_{\perp i}$ and r_i are the shape parameters of the sources. The pointing errors, l_j and m_j enter the Jones matrices, along with the beam width, bw.

5.6.3.1 Results

Figure 5.23 shows the 1d marginalised posterior for three of the parameters for the "QMC shape" dataset, demonstrating that we are able to accurately determine the true value of each parameter

to within the estimated uncertainty. All the histograms for this dataset are in Appendix A. With this dataset, we demonstrate we can not only estimate the pointing errors and the fluxes, but also estimate all the shape parameters and distinguish with statistical significance, between a point and an extended source.

An observation of the histograms will reveal the occasional bimodal posterior in a few of the point source shape parameters. This is simply due to a degeneracy in the setup of the problem. For example, a pointing angle of $-45°$ and $135°$ produces exactly the same Gaussian, but one will have a negative $l_\perp$ and the other positive. Hence when estimating the shape parameters which are expected to be close to zero, it is not surprising to observe a double peak as the MCMC prefers a slightly extended source (due to random noise in the data).

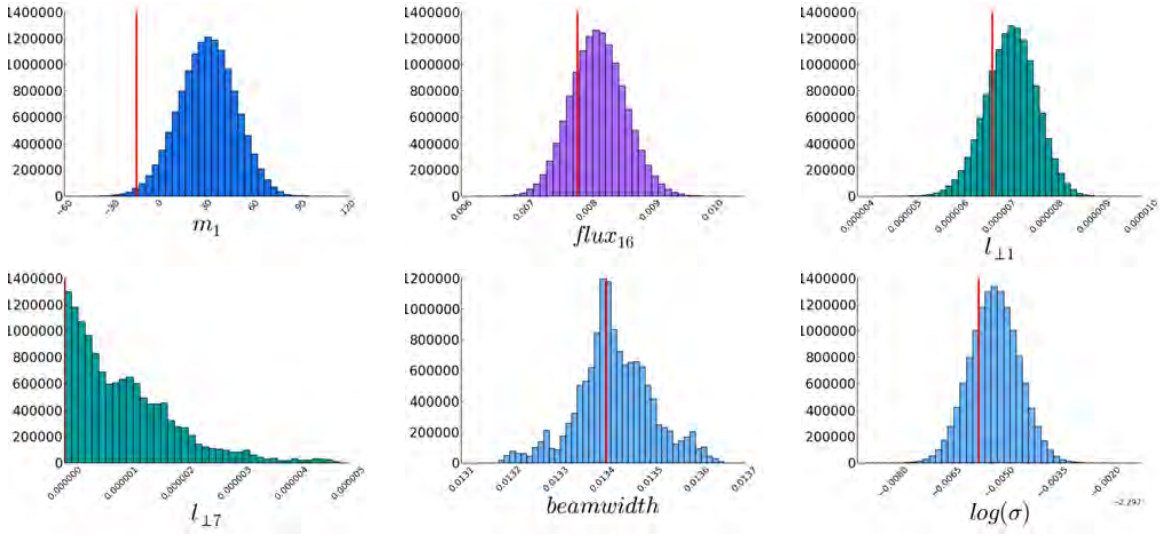


FIGURE 5.23: Selected 1d marginalised posteriors from the MCMC analysis of the "QMC shape" dataset. The true value is given by the red vertical line. Parameters are (reading from left to right, top to bottom) m_1: the pointing error in the declination direction for the first antenna, $flux_{17}$: the flux of one of the point sources, $l_\perp 1$: the projection of the major axis of the extended source on the RA-direction, $l\text{proj}_7$: the projection of the major axis of a non-extended source on the RA-direction, *beamwidth*: parameter c in Eq.(5.32) and $log(\sigma)$: the log of the standard deviation of the visibility noise.

We compared our parameter estimates now including the additional extended sources parameters with the CLEAN+SE estimates of Section 5.6.2 (Figure 5.24). The true fluxes are consistent with the BIRO estimates, the CLEAN+SE results are obviously the same as in Figure **??** (see also Table A.3 in Appendix A). Crucially, CLEAN+SE fails, at high confidence, to detect the extended source and also mis-classifies a few point sources as extended sources. BIRO on the other hand, is given complete freedom in the model: under the assumption that we do not know whether a given source is a point source or actually extended, we treat them all as extended sources and fit for their parameters. The result is that the parameters for the extended source

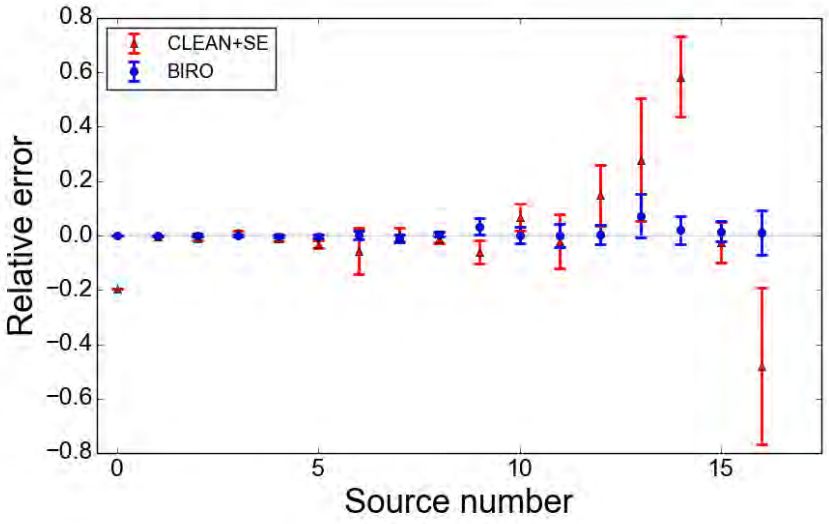


FIGURE 5.24: Comparison between BIRO and the results from CLEANing and then applying a source extraction algorithm to the data (CLEAN+SE). For each of the 17 sources in the dataset, we subtract the estimated flux from the true flux and divide the result by the true flux to obtain and plot the relative error (y-axis). We similarly scale the error bars for both BIRO and CLEAN+SE by dividing them by the true flux.

are non-zero at high confidence, while all the other sources are found to be consistent with being point sources. Thus, unlike CLEAN, we do not have to make the fundamental assumption that everything is a point source. This could be of particular interest when, for example, analysing a weak lensing field with the SKA [153, 154] and fitting for the ellipticity of sources.

We plot the covariance matrix for the "QMC" dataset with all shape parameters allowed to vary (Figure 5.25), using the same code used to produce Figure 5.15 and Figure 5.21. The caption of the figure explains what the colours and shapes of the ellipses mean. Again there are strong correlations between instrumental and scientific parameters.

The interaction between pointing errors and fluxes is obviously highly complex from looking at Figure 5.25. The beam width is highly correlated with almost all the other parameters, implying that modeling and understanding the beam is of crucial importance. Some individual shape parameters are completely uncorrelated with most other parameters, while others correlate or anti-correlate strongly with the pointing errors and/or the estimates of the fluxes, depending most likely on where the source is in the sky relative to the beam. This covariance matrix and further controlled simulations, could be used to untangle and understand these correlations.

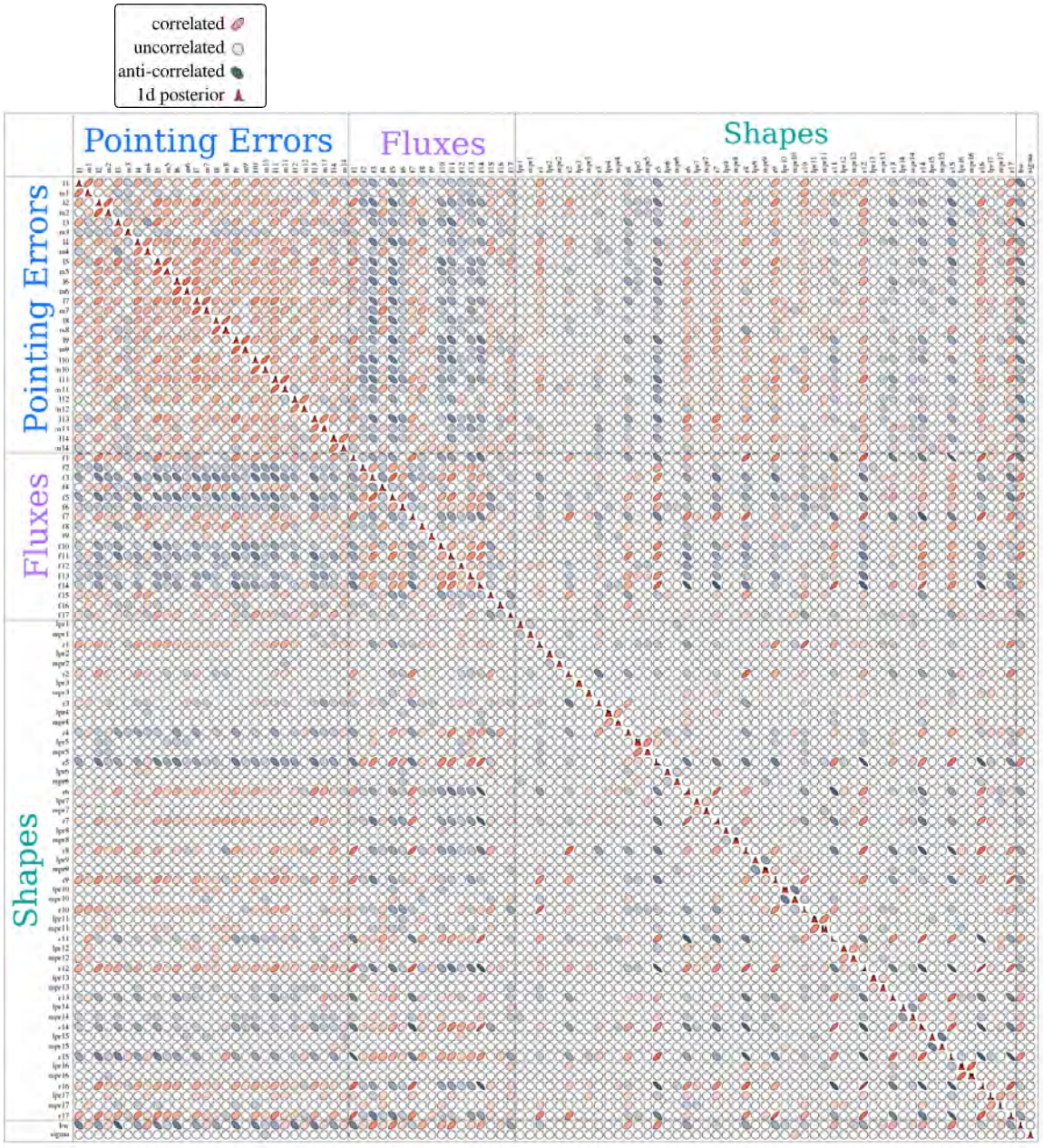


FIGURE 5.25: Covariance matrix between parameters for the "QMC shapw" simulated dataset. The parameters are listed on each axis (pointing errors for each antenna first, followed by the fluxes, the shape parameters, the beam width and the log of the noise) with the correlations between them represented by a coloured ellipse. Highly correlated parameters are red with a thin, angled ellipse, whereas anti-correlated parameters have a dark blue ellipse, angled to the left. The diagonal shows the 1d marginalised posterior for each parameter (the same as the above histograms).

5.7 Discussion

We have introduced the technique Bayesian Inference for Radio Observations (BIRO), a Bayesian approach to the deconvolution problem of radio interferometry. Instead of making an image and then performing source extraction, BIRO uses MCMC to fit models directly to the data and obtain the posterior for the parameters of interest, as well as nuisance parameters.

We focused on the relationship between scientific and instrumental parameters. In general, the scientific parameters we considered were fluxes of sources and the shape parameters of extended sources. We also included in our simulations pointing errors, beam width and visibility noise as instrumental parameters. We then tested BIRO on three simulated datasets of increasing complexity by first considering only point sources and constant pointing errors, then adding an extended source and lastly allowing the shape parameters of all sources to vary.

Our results prove that BIRO as a technique works. It was found that all parameter estimates were consistent within their error bars with the true values. We were able to, for the first time, estimate the uncertainty on pointing errors for a typical WSRT dataset. We were also able to place believable error bars on our scientific parameters which included the effects of all the other parameters. Additionally, we could accurately estimate the noise of the visibilities directly from the data.

Another important product of BIRO is the covariance matrix between parameters. For example, we found for WSRT the pointing errors in each direction were highly correlated. The fluxes were also highly correlated with each other. Most important to this work is the correlations between instrumental and scientific parameters. These are generally ignored in current techniques. Our work shows these correlations are complicated and non-negligable. BIRO effortlessly incorporates the effects of the correlations in the estimates of the marginalised uncertainties on the individual parameters, as well as providing a way to study these correlations in the form of the covariance matrix.

BIRO's main competitor is CLEAN, the most popular deconvolution algorithm. CLEAN produces an image from the data, which a source extraction algorithm is then applied to in order to recover the parameters of the sources in the image. This process is an optimisation algorithm, which purely finds the best fitting parameters it can. It is also a two step process, with assumptions in both steps. This is in contrast to BIRO which, in one step, extracts both estimates and uncertainty information directly from the data without making an image.

To compare the two algorithms for each dataset, we performed a CLEAN+source extractor (CLEAN+SE) analysis. Without any information about the pointing errors, CLEAN+SE produces catastrophic biases in the estimated fluxes of the sources, up to 76% error in some cases. Worse than this is the fact that the uncertainties placed on these fluxes are so small that this translates to biases of over 100-σ. Not only does this traditional analysis ignore the instrumental effects that may be affecting even current observations, it cannot include any uncertainty information about this assumption.

BIRO is highly versatile, since anything that can be modeled with MeqTrees can be estimated using BIRO. We are currently involved in a project to use BIRO for source separation: using the Bayesian evidence (see Section 3.4.1) to determine whether what appears as an ambiguous "blob" to the eye is statistically more likely to be a single extended source or two individual sources. BIRO could also be applied to the study of variable radio sources, the structure of objects such as jets from AGN and artifacts seen in radio images.

BIRO is particularly applicable to the SKA. To achieve the full potential of the sensitivity of the SKA, systematic errors will have to be controlled to unprecedented levels. Current techniques either do not deal with these errors or are *ad hoc* in their approach. The RIME allows a rigorous mathematical approach to interferometry and BIRO as a technique can leverage it to obtain the maximum information from the SKA datasets.

Of course, the SKA is a very different scale of telescope to what we have been working with. The SKA will have of the order of thousands of antennas and millions of baselines. While this implies the amount of data will far outnumber the parameters, it still means there will be thousands of highly correlated, nuisance parameters before we even begin to consider the scientific parameters. The computational complexity of MeqTrees scales as the number of antennas squared, making the likelihood difficult to compute for thousands of antennas. However, it is also a highly parallelisable operation making it computationally feasible on a cluster or GPU.

While MCMC with large numbers of parameters with current techniques is possible [155], we may like a sampler that can also deliver the evidence for the model to allow model comparison, such as Hamiltonian Monte Carlo [156, 157] with thermodynamic integration. An affine invariant sampler [158, 159] may also be appropriate for the large dimensionality and high computation cost of the likelihood since large-scale parallelisation is possible with such samplers. There will be much opportunity for development in the realm of high dimensional MCMC samplers for BIRO as the SKA draws closer to completion.

Chapter 6

Conclusions

Cosmology and astronomy have entered the era of precision experiments and big data. Previously neglected systematic effects can no longer be ignored as the constraints on scientific parameters get better and better. As long as these systematics are ignored, biased science results and incorrect estimates of the uncertainties are almost inevitable.

In this thesis, we have tackled this issue in two very different problems but with the same set of tools: Bayesian statistics (see Chapter 3). With these tools, we are able to quantify and propagate uncertainties consistently within the models. Although Bayesian techniques tend to be more computationally expensive than traditional techniques, we cannot escape from the need for more sophisticated treatment of uncertainties if we are to achieve the maximum potential of next generation surveys such as the SKA. Fortunately, it seems the continued growth in computing power will allow the application of Bayesian statistics to even these massive datasets.

In Chapter 4, we demonstrated the applicability of Bayesian statistics in dealing with the contamination from non-Type Ia (nIa) supernovae in photometric supernova data. This contamination will be inevitable in future surveys such as LSST because of the flood of high quality supernovae candidates, for which spectroscopic typing will not be possible. Using this contaminated dataset can result in large biases in the estimates of the cosmological parameters. Bayesian Estimation for Multiple Species (BEAMS), is a Gaussian mixture model formalism designed to incorporate this type uncertainty into the analysis to produce unbiased results. We extended BEAMS to deal with correlated supernovae data, since the original formalism assumed the data were uncorrelated. Correlations in the supernova data can arise from peculiar velocities, filter errors, observational conditions etc.

The analytic correlated BEAMS likelihood is an order 2^N calculation, which is computationally impossible. To solve this problem, we developed two alternatives to the analytic form. The first, a perturbative approximation, failed due to its poor approximation of the true likelihood. The second however, a numerical marginalisation over the types of the supernovae, succeeded in accurately estimating the cosmological parameters for simulated datasets using three different types of covariance matrix to correlate the data. This numerical marginalisation required the sampling of over 200 parameters with MCMC (see Chapter 3). To ensure that our results were consistent and correct, we ran 10 000 MCMC chains (5000 each for the uncorrelated and correlated assumption in the likelihood), resulting in a total of over a billion MCMC steps taken. We found that for correlated datasets, our numerical marginalisation form of the BEAMS likelihood was able to provide good coverage (showing our credible interval estimates are accurate) while the original BEAMS formalism, assuming uncorrelated data, had very poor coverage.

In Chapter 5, we also applied Bayesian statistics to the very different field of radio astronomy, specifically in the deconvolution problem of radio interferometry, where there is always ambiguity due to missing information from the gaps between individual antennae in the interferometer. The goal is to obtain, from raw radio observations (the visibilities), scientific parameters such as a catalogue of source fluxes. Traditionally, this problem is solved with a combination of the algorithm CLEAN and source extraction (we refer to this combination as CLEAN+SE). This is purely an optimisation algorithm, first finding the best image it can fit to the data, then finding the best fitting parameters. CLEAN+SE has serious drawbacks including a lack of uncertainty information and a way to deal with correlations between instrumental and scientific parameters. In this chapter, we proposed a new technique, Bayesian Inference for Radio Observations (BIRO), which uses MCMC and model fitting with MeqTrees to determine the best fitting parameters, their uncertainties and correlations (i.e. the full posterior).

We proved the effectiveness of BIRO over CLEAN+SE by generating three datasets (simulating WSRT data) of increasing complexity. We included instrumental parameters such as pointing errors in the antennae and the beam width of the primary beam. Our science parameters included the fluxes of the sources in the field and the parameters of extended sources. We were also able to recover in each case the noise on the visibilities. CLEAN+SE produced highly biased results for the fluxes and the shape parameters of the sources, vastly underestimating the error bars, for almost every source. BIRO, however, was able to accurately estimate every parameter for every dataset, even for our most complex dataset which had 98 highly correlated parameters, and provide a believable uncertainty for them. This is, to our knowledge, the

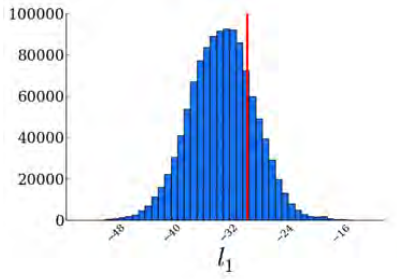
first time anyone has been able to determine a reliable uncertainty for a pointing error, or the uncertainty for the flux of source which marginalises over the effects of all known instrumental effects. BIRO was also able to reliably distinguish between a Gaussian extend source and point sources in the field, whereas CLEAN+SE returned poor estimates and highly biased estimates of the shape parameters and source classifications. As telescopes become more and more powerful, new techniques such as BIRO will be required to unlock their true potential.

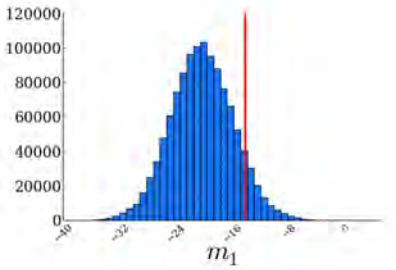
Appendix A

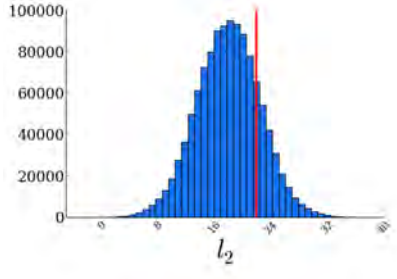
Additional BIRO Results

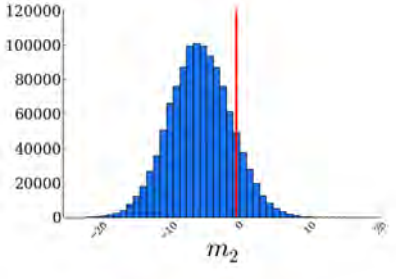
A.1 Grid mock data

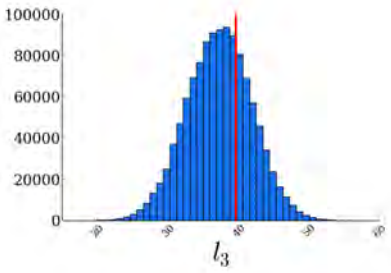
Below are the one-dimensional marginalised posteriors for all the parameters of this simulated dataset, illustrating that the true, input parameters (indicated in each case by a vertical line) are well recovered in all cases. For the pointing errors, l_i indicates the pointing error in right ascension direction for the $i'th$ antenna and similarly, m_i indicates the pointing error in the declination direction for the $i'th$ antenna.

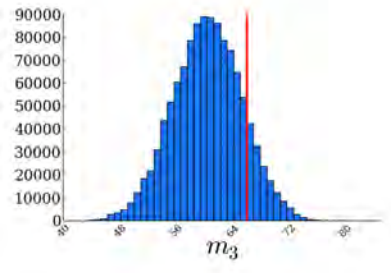


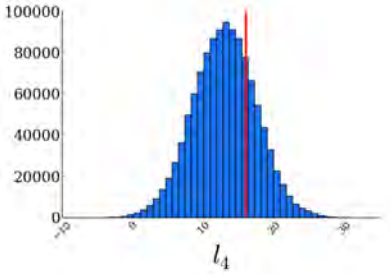


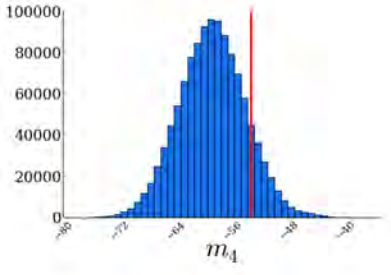


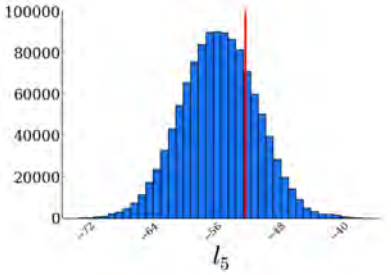


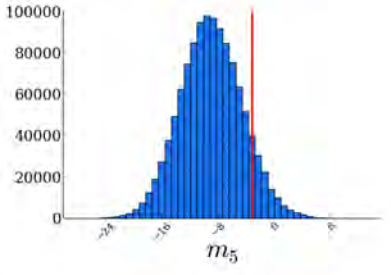

100000
80000
60000
40000
20000
0
l_3

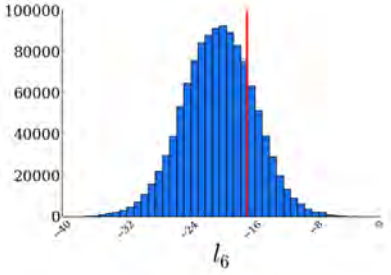

90000
80000
70000
60000
50000
40000
30000
20000
10000
0
m_3

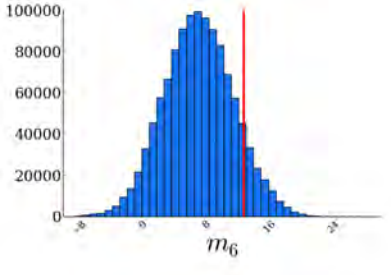

100000
80000
60000
40000
20000
0
l_4

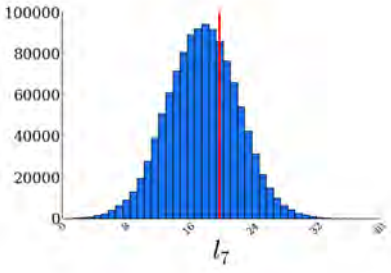

100000
80000
60000
40000
20000
0
m_4

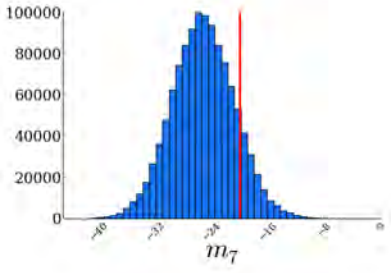

100000
80000
60000
40000
20000
0
l_5

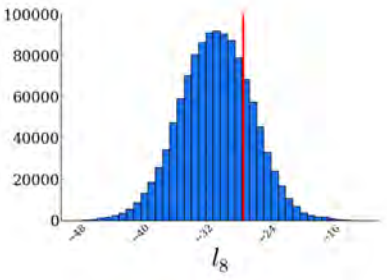

100000
80000
60000
40000
20000
0
m_5

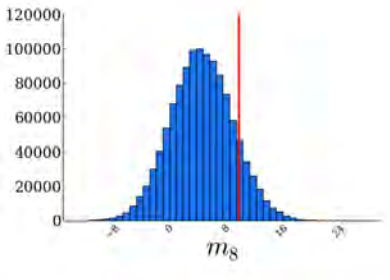

100000
80000
60000
40000
20000
0
l_6

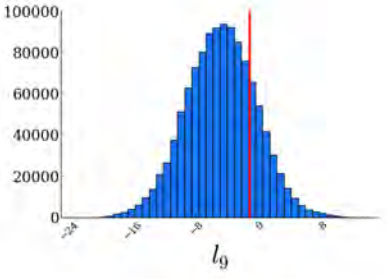

100000
80000
60000
40000
20000
0
m_6

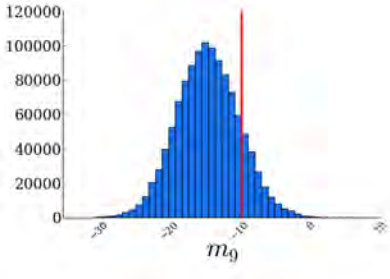

100000
80000
60000
40000
20000
0
l_7

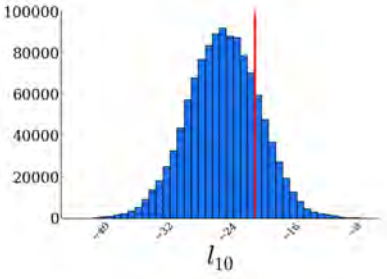

100000
80000
60000
40000
20000
0
m_7

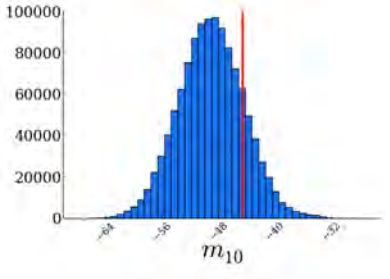

100000
80000
60000
40000
20000
0
l_8

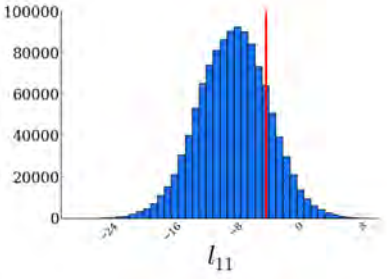

120000
100000
80000
60000
40000
20000
0
m_8

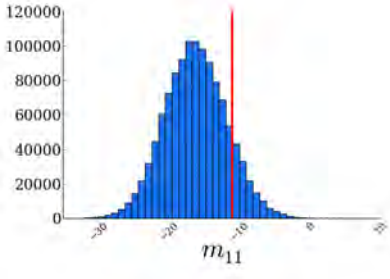

100000
80000
60000
40000
20000
0
l_9

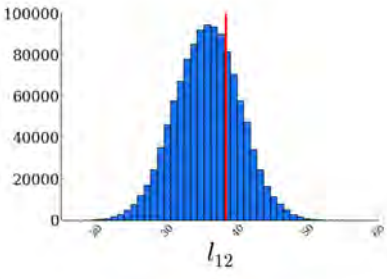

120000
100000
80000
60000
40000
20000
0
m_9

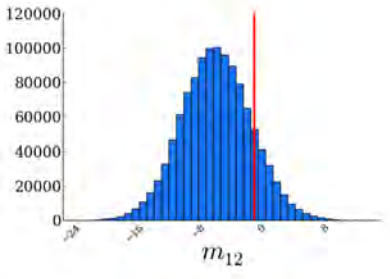

100000
80000
60000
40000
20000
0
l_{10}

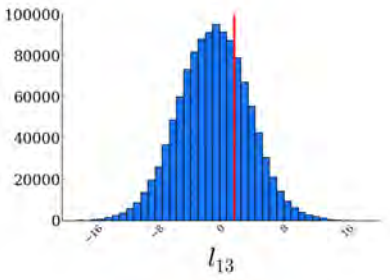

100000
80000
60000
40000
20000
0
m_{10}

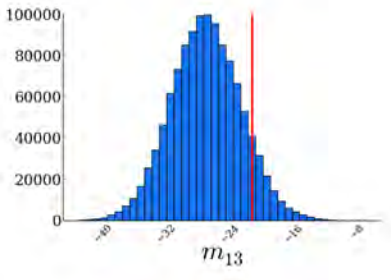

100000
80000
60000
40000
20000
0
l_{11}

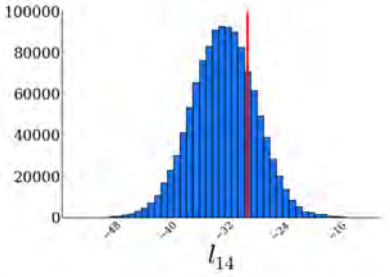

120000
100000
80000
60000
40000
20000
0
m_{11}

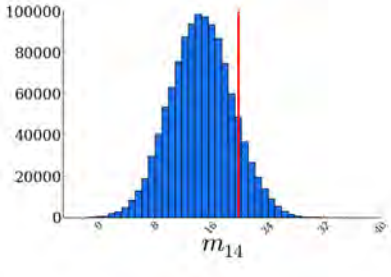

100000
80000
60000
40000
20000
0
l_{12}

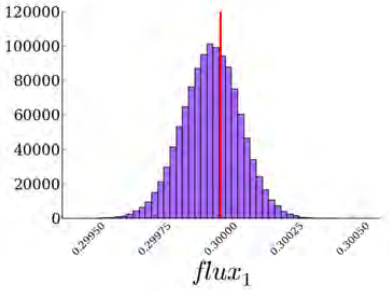

120000
100000
80000
60000
40000
20000
0
m_{12}

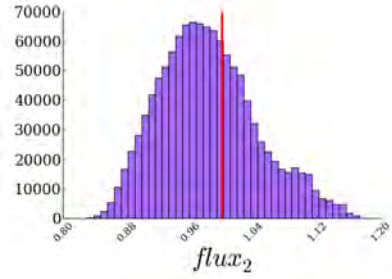

100000
80000
60000
40000
20000
0
−16
−8
0
8
16
l_{13}

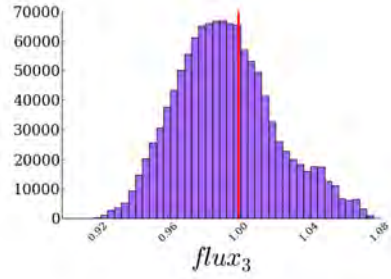

100000
80000
60000
40000
20000
0
−40
−32
−24
−16
−8
m_{13}

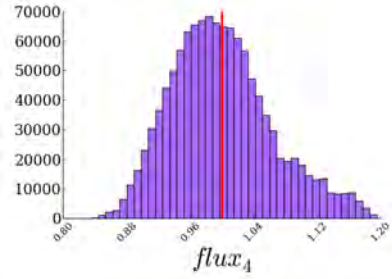

100000
80000
60000
40000
20000
0
−48
−40
−32
−24
−16
l_{14}

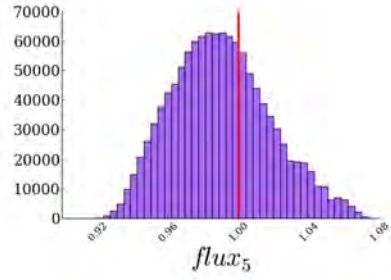

100000
80000
60000
40000
20000
0
0
8
16
24
32
40
m_{14}

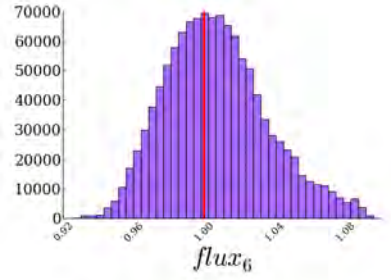

120000
100000
80000
60000
40000
20000
0
0.29950
0.29975
0.30000
0.30025
0.30050
$flux_1$

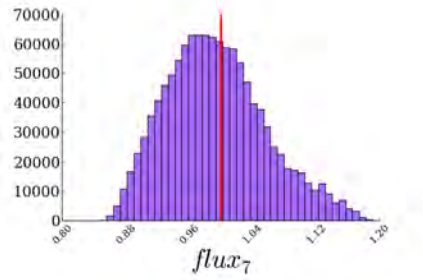

70000
60000
50000
40000
30000
20000
10000
0
0.80
0.88
0.96
1.04
1.12
1.20
$flux_2$

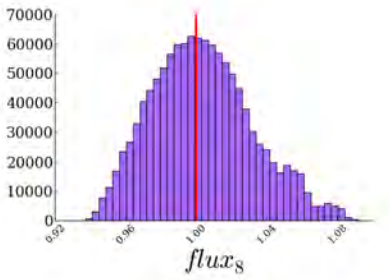

70000
60000
50000
40000
30000
20000
10000
0
0.92
0.96
1.00
1.04
1.08
$flux_3$

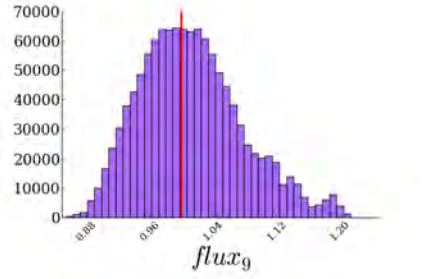

70000
60000
50000
40000
30000
20000
10000
0
0.80
0.88
0.96
1.04
1.12
1.20
$flux_4$

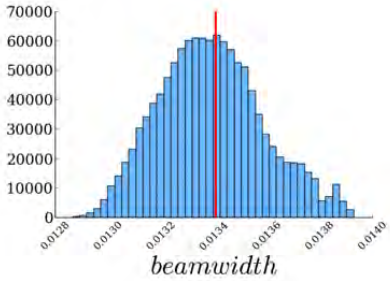

70000
60000
50000
40000
30000
20000
10000
0
0.92
0.96
1.00
1.04
1.08
$flux_5$

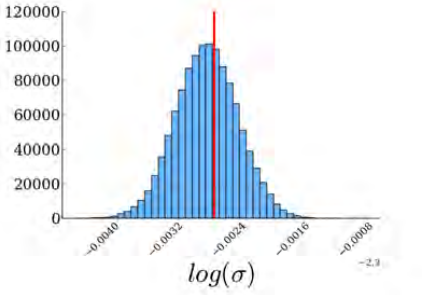

70000
60000
50000
40000
30000
20000
10000
0
0.92
0.96
1.00
1.04
1.08
$flux_6$


70000
60000
50000
40000
30000
20000
10000
0
0.80
0.88
0.96
1.04
1.12
1.20
$flux_7$


70000
60000
50000
40000
30000
20000
10000
0
0.92
0.96
1.00
1.04
1.08
$flux_8$


70000
60000
50000
40000
30000
20000
10000
0
0.88
0.96
1.04
1.12
1.20
$flux_9$


70000
60000
50000
40000
30000
20000
10000
0
0.0128
0.0130
0.0132
0.0134
0.0136
0.0138
0.0140
$beamwidth$

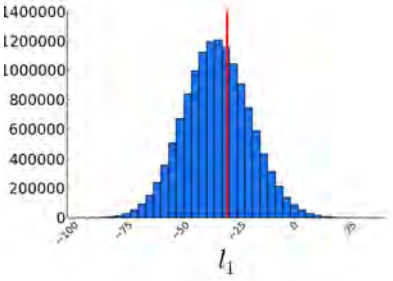

120000
100000
80000
60000
40000
20000
0
−0.0040
−0.0032
−0.0024
−0.0016
−0.0008
$log(\sigma)$

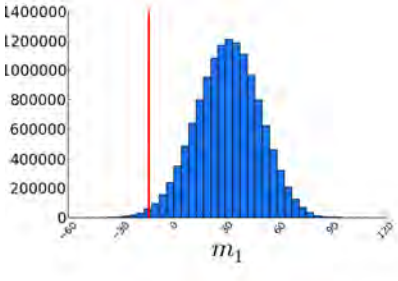
Name	RA (°)	Dec (°)	True Flux Density (Jy)	CLN Flux Density (Jy)	BIRO Flux Density (Jy)
S0	0.000	60.000	0.3000	$0.1609 \pm 0.0003\,(512.8\sigma)$	$0.3000^{+0.0001}_{-0.0001}\,(0.4\sigma)$
S1	359.340	59.665	1.0000	$0.9813 \pm 0.0017\,(10.8\sigma)$	$0.9808^{+0.0602}_{-0.0602}\,(0.3\sigma)$
S2	359.333	59.998	1.0000	$0.9941 \pm 0.0007\,(9.1\sigma)$	$0.9913^{+0.0278}_{-0.0278}\,(0.3\sigma)$
S3	359.327	60.332	1.0000	$0.7643 \pm 0.0018\,(133.6\sigma)$	$1.0021^{+0.0630}_{-0.0630}\,(0.0\sigma)$
S4	0.000	59.667	1.0000	$0.9912 \pm 0.0007\,(13.5\sigma)$	$0.9910^{+0.0270}_{-0.0270}\,(0.3\sigma)$
S5	0.000	60.333	1.0000	$0.7874 \pm 0.0007\,(322.5\sigma)$	$1.0089^{+0.0298}_{-0.0298}\,(0.3\sigma)$
S6	0.660	59.665	1.0000	$0.8516 \pm 0.0017\,(85.5\sigma)$	$0.9984^{+0.0603}_{-0.0603}\,(0.0\sigma)$
S7	0.667	59.998	1.0000	$0.2348 \pm 0.0007\,(1163.4\sigma)$	$1.0050^{+0.0310}_{-0.0271}\,(0.2\sigma)$
S8	0.673	60.332	1.0000	$0.3990 \pm 0.0018\,(336.1\sigma)$	$1.0121^{+0.0628}_{-0.0628}\,(0.2\sigma)$
Extra source	0.703	59.6581		0.0152 ± 0.0016	
Extra source	0.660	59.6637		0.1264 ± 0.0018	
Extra source	0.666	59.9960		0.1709 ± 0.0007	
Extra source	0.669	59.9991		0.4340 ± 0.0007	
Extra source	0.665	60.0017		0.1424 ± 0.0007	
Extra source	0.674	60.3315		0.5307 ± 0.0018	
Extra source	0.673	60.3373		0.0441 ± 0.0018	
Extra source	0.696	60.3301		0.0035 ± 0.0019	
Extra source	0.694	60.3459		0.0080 ± 0.0019	
Extra source	0.614	59.8209		0.0087 ± 0.0005	
Extra source	0.434	60.1696		0.0051 ± 0.0004	
Extra source	360.000	59.9990		0.1243 ± 0.0003	
Extra source	359.999	59.9960		0.0138 ± 0.0003	
Extra source	360.000	60.3314		0.1579 ± 0.0007	
Extra source	0.001	60.3336		0.0398 ± 0.0007	
Extra source	359.353	60.0178		0.0015 ± 0.0006	
Extra source	359.319	60.0007		0.0039 ± 0.0006	
Extra source	359.340	59.9918		0.0013 ± 0.0006	
Extra source	359.324	60.3316		0.1981 ± 0.0018	
Extra source	359.329	60.3351		0.0269 ± 0.0015	
Extra source	359.317	59.6421		0.0810 ± 0.0009	
Extra source	359.162	59.9947		0.0037 ± 0.0011	

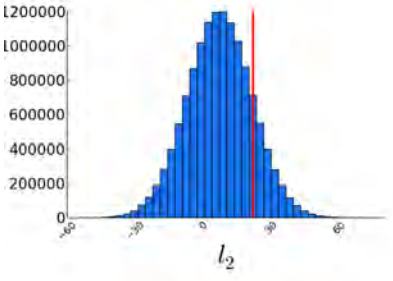
TABLE A.1: Comparison between the CLEAN+source extraction results (shortened to CLN) and the BIRO results for the "grid" mock dataset.

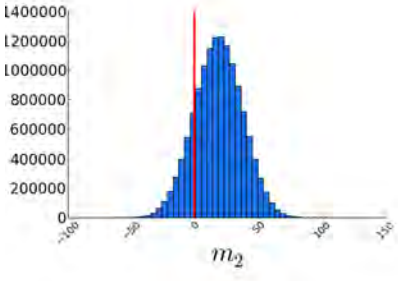
A.2 QMC mock data

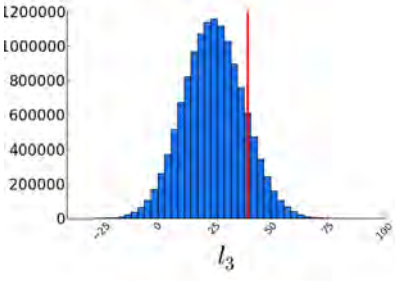
Below are the one-dimensional marginalised posteriors for all the parameters of this simulated dataset, illustrating that the true, input parameters (indicated in each case by a vertical line) are well recovered in all cases. For the pointing errors, l_i indicates the pointing error in right ascension direction for the $i'th$ antenna and similarly, m_i indicates the pointing error in the declination direction for the $i'th$ antenna. The shape parameters for the extended source are given by $l$proj, $m$proj and $r = e\text{min}/e\text{maj}$ (see Figure 5.8).

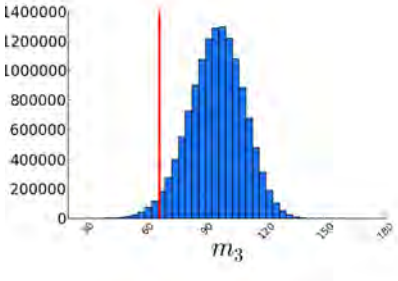

1400000
1200000
1000000
800000
600000
400000
200000
0
l_1

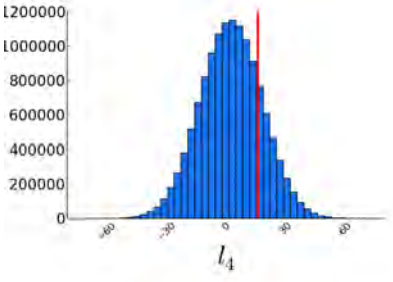

1400000
1200000
1000000
800000
600000
400000
200000
0
m_1

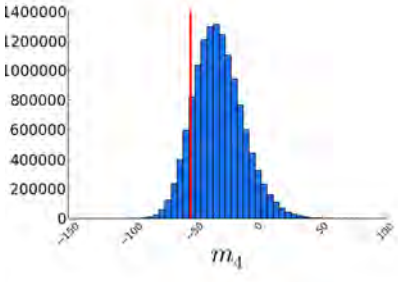

1200000
1000000
800000
600000
400000
200000
0
l_2

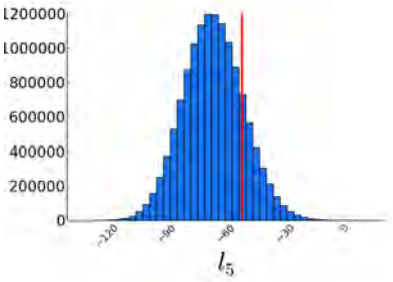

1400000
1200000
1000000
800000
600000
400000
200000
0
m_2

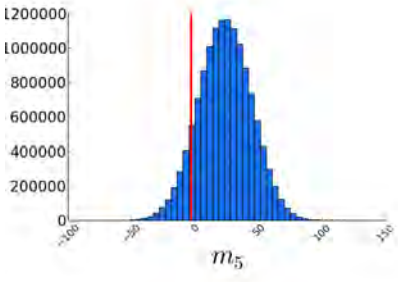

1200000
1000000
800000
600000
400000
200000
0
l_3

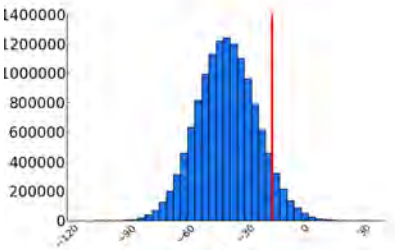

1400000
1200000
1000000
800000
600000
400000
200000
0
m_3

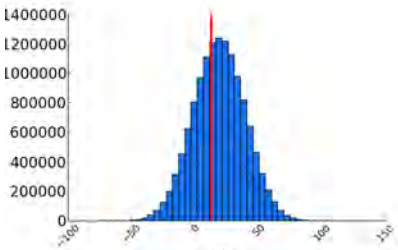

1200000
1000000
800000
600000
400000
200000
0
l_4

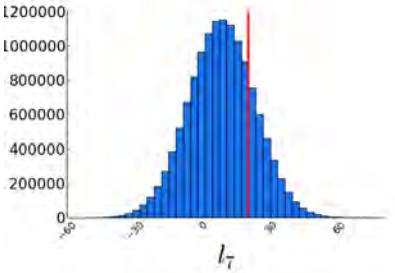

1400000
1200000
1000000
800000
600000
400000
200000
0
m_4

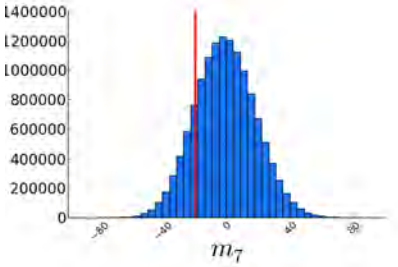

1200000
1000000
800000
600000
400000
200000
0
l_5

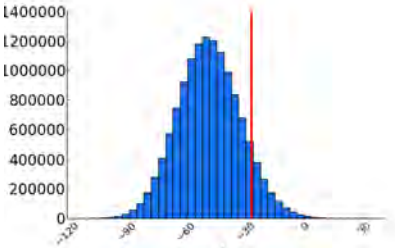

1200000
1000000
800000
600000
400000
200000
0
m_5

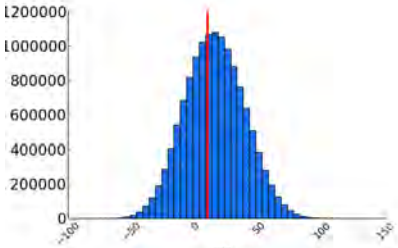

1400000
1200000
1000000
800000
600000
400000
200000
0
l_6

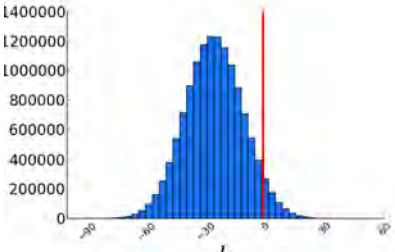

1400000
1200000
1000000
800000
600000
400000
200000
0
m_6

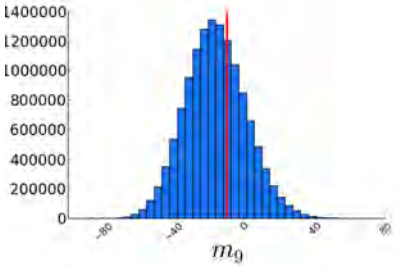

1200000
1000000
800000
600000
400000
200000
0
l_7

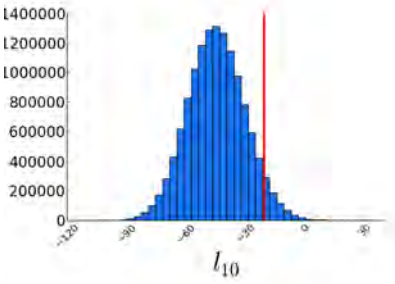

1400000
1200000
1000000
800000
600000
400000
200000
0
m_7

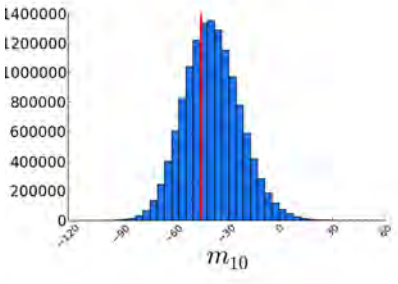

1400000
1200000
1000000
800000
600000
400000
200000
0
l_8

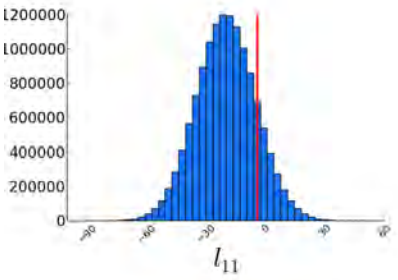

1200000
1000000
800000
600000
400000
200000
0
m_8

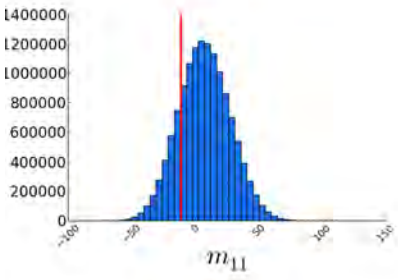

1400000
1200000
1000000
800000
600000
400000
200000
0
l_9

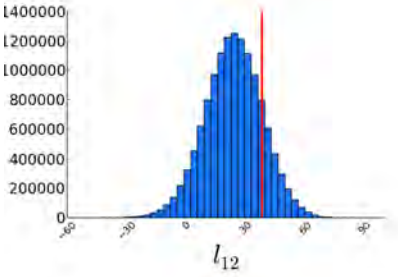

1400000
1200000
1000000
800000
600000
400000
200000
0
m_9

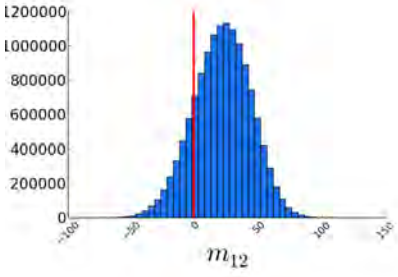

1400000
1200000
1000000
800000
600000
400000
200000
0
l_{10}

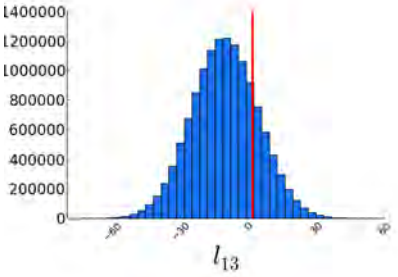

1400000
1200000
1000000
800000
600000
400000
200000
0
m_{10}

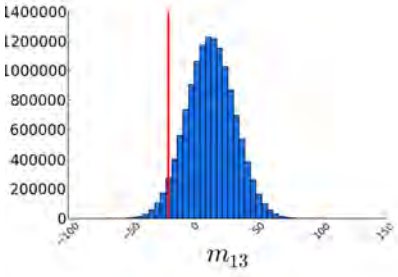

l_{11}

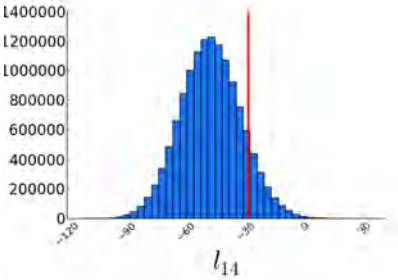

m_{11}

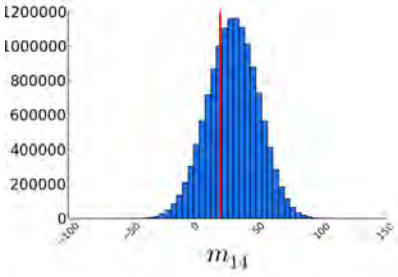

l_{12}

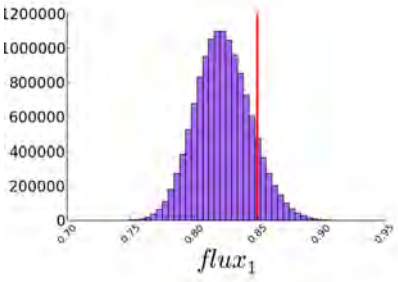

m_{12}

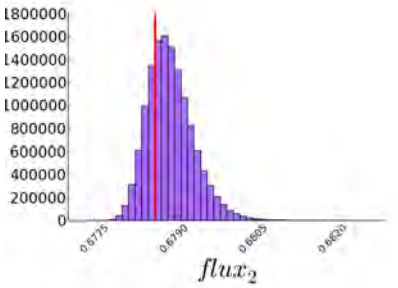

l_{13}

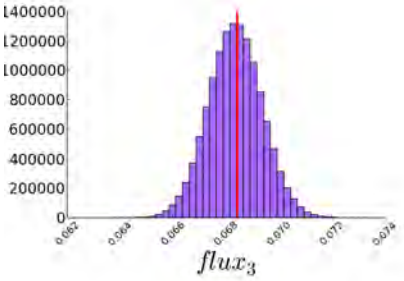

m_{13}

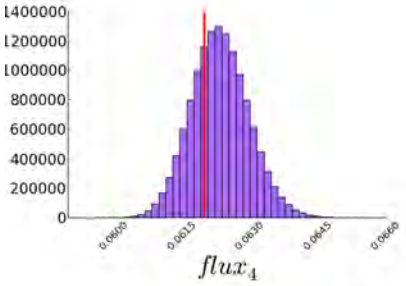

l_{14}

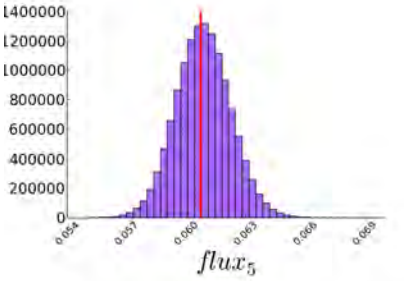

m_{14}

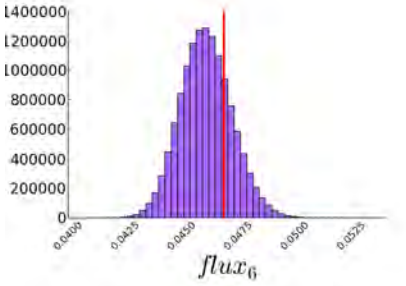

$flux_1$

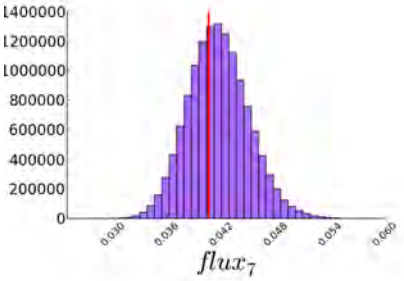

$flux_2$

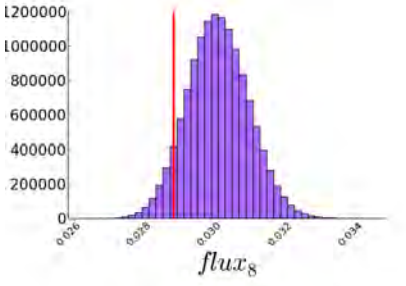

1400000
1200000
1000000
800000
600000
400000
200000
0
0.062
0.064
0.066
0.068
0.070
0.072
0.074
$flux_3$

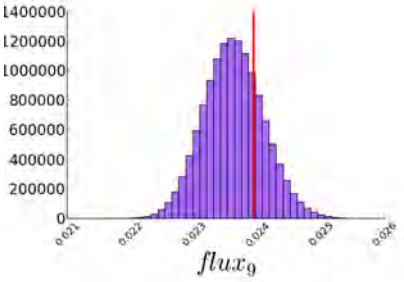

1400000
1200000
1000000
800000
600000
400000
200000
0
0.0600
0.0615
0.0630
0.0645
0.0660
$flux_4$

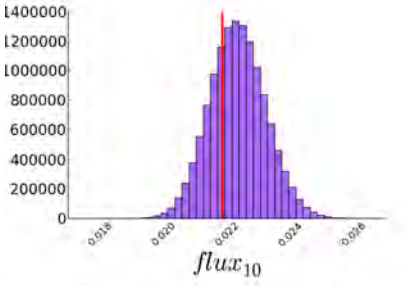

1400000
1200000
1000000
800000
600000
400000
200000
0
0.054
0.057
0.060
0.063
0.066
0.069
$flux_5$

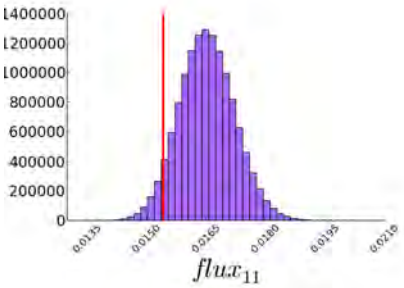

1400000
1200000
1000000
800000
600000
400000
200000
0
0.0400
0.0425
0.0450
0.0475
0.0500
0.0525
$flux_6$

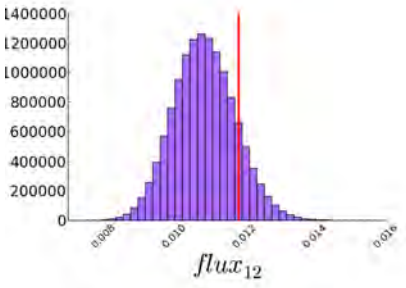

1400000
1200000
1000000
800000
600000
400000
200000
0
0.030
0.036
0.042
0.048
0.054
0.060
$flux_7$

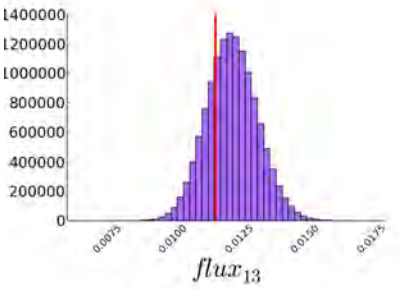

1200000
1000000
800000
600000
400000
200000
0
0.026
0.028
0.030
0.032
0.034
$flux_8$

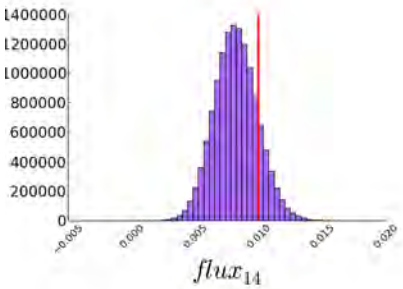

1400000
1200000
1000000
800000
600000
400000
200000
0
0.021
0.022
0.023
0.024
0.025
0.026
$flux_9$

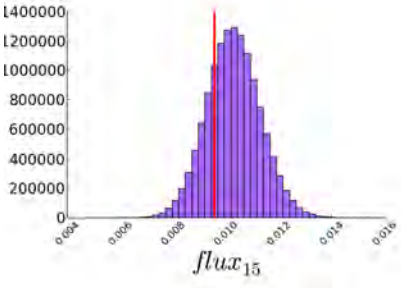

1400000
1200000
1000000
800000
600000
400000
200000
0
0.018
0.020
0.022
0.024
0.026
$flux_{10}$

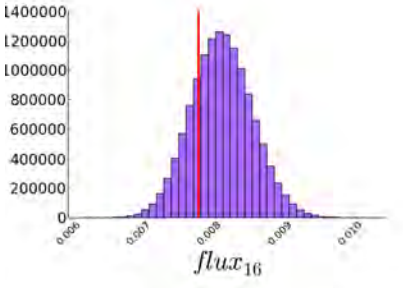

1400000
1200000
1000000
800000
600000
400000
200000
0
0.0135
0.0150
0.0165
0.0180
0.0195
0.0210
$flux_{11}$

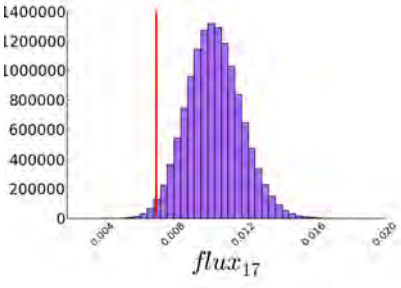

1400000
1200000
1000000
800000
600000
400000
200000
0
0.008
0.010
0.012
0.014
0.016
$flux_{12}$

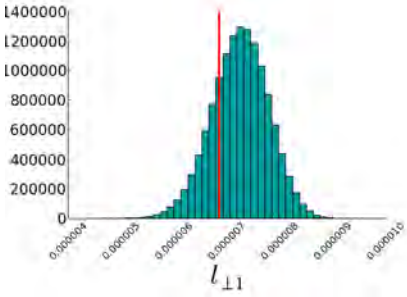

$flux_{13}$

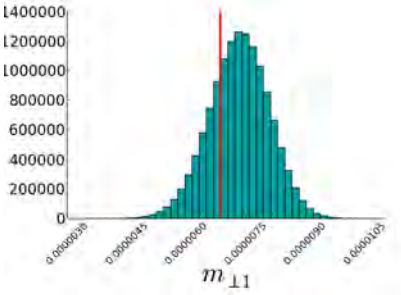

$flux_{14}$

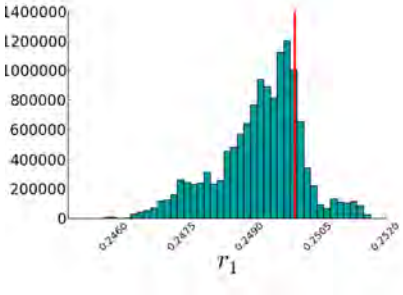

$flux_{15}$

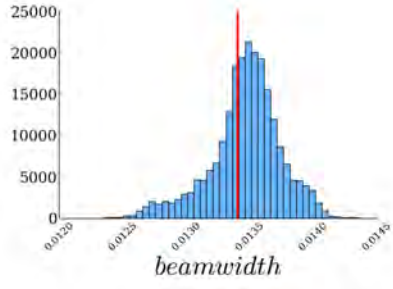

$flux_{16}$

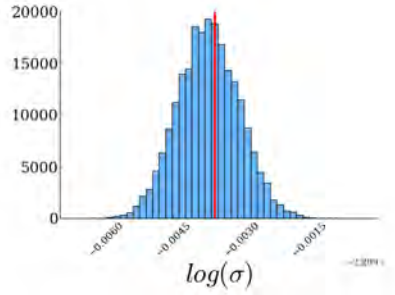

$flux_{17}$


$l_{\perp 1}$


$m_{\perp 1}$


r_1


$beamwidth$

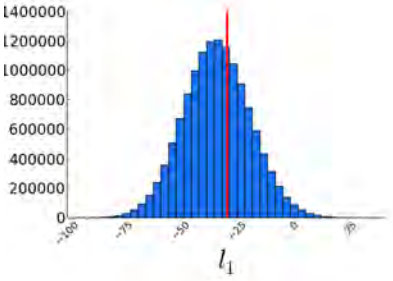

$log(\sigma)$

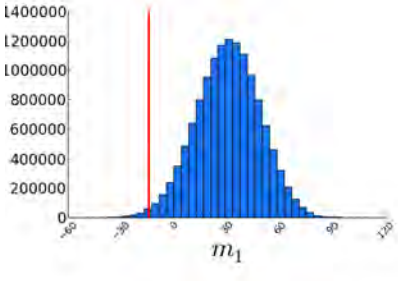
Name	RA (°)	Dec (°)	True Flux Density (Jy)	CLN Flux Density (Jy)	BIRO Flux Density (Jy)
B292	44.782	50.894	0.8500	$0.6837 \pm 0.0009\,(189.7\sigma)$	$0.8385^{+0.0261}_{-0.0209}\,(0.5\sigma)$
A240	45.303	50.752	0.6786	$0.6763 \pm 0.0003\,(7.6\sigma)$	$0.6793^{+0.0005}_{-0.0005}\,(1.4\sigma)$
C051	45.574	50.904	0.0684	$0.0677 \pm 0.0005\,(1.5\sigma)$	$0.0700^{+0.0019}_{-0.0015}\,(0.9\sigma)$
D221	45.150	50.622	0.0620	$0.0626 \pm 0.0004\,(1.7\sigma)$	$0.0611^{+0.0007}_{-0.0007}\,(1.3\sigma)$
E041	45.655	50.974	0.0607	$0.0601 \pm 0.0007\,(1.0\sigma)$	$0.0626^{+0.0028}_{-0.0023}\,(0.7\sigma)$
F321	45.035	51.019	0.0469	$0.0454 \pm 0.0007\,(2.1\sigma)$	$0.0465^{+0.0014}_{-0.0011}\,(0.3\sigma)$
G223	44.740	50.358	0.0406	$0.0383 \pm 0.0035\,(0.7\sigma)$	$0.0382^{+0.0048}_{-0.0040}\,(0.6\sigma)$
H182	45.339	50.432	0.0290	$0.0291 \pm 0.0007\,(0.2\sigma)$	$0.0282^{+0.0012}_{-0.0012}\,(0.6\sigma)$
I281	44.953	50.839	0.0239	$0.0237 \pm 0.0005\,(0.5\sigma)$	$0.0238^{+0.0005}_{-0.0005}\,(0.4\sigma)$
J091	45.885	50.760	0.0219	$0.0205 \pm 0.0009\,(1.4\sigma)$	$0.0239^{+0.0016}_{-0.0016}\,(1.3\sigma)$
K151	45.662	50.510	0.0154	$0.0164 \pm 0.0008\,(1.4\sigma)$	$0.0153^{+0.0009}_{-0.0008}\,(0.1\sigma)$
L042	45.706	51.058	0.0118	$0.0116 \pm 0.0012\,(0.2\sigma)$	$0.0122^{+0.0010}_{-0.0010}\,(0.4\sigma)$
M072	45.919	50.897	0.0116	$0.0133 \pm 0.0013\,(1.3\sigma)$	$0.0122^{+0.0014}_{-0.0012}\,(0.5\sigma)$
N042	45.851	51.092	0.0100	$0.0128 \pm 0.0023\,(1.2\sigma)$	$0.0144^{+0.0023}_{-0.0023}\,(1.9\sigma)$
O322	44.973	51.139	0.0096	$0.0152 \pm 0.0014\,(3.9\sigma)$	$0.0102^{+0.0010}_{-0.0010}\,(0.6\sigma)$
P111	45.744	50.719	0.0078	$0.0077 \pm 0.0006\,(0.3\sigma)$	$0.0080^{+0.0005}_{-0.0005}\,(0.4\sigma)$
Q182	45.434	50.288	0.0070	$0.0037 \pm 0.0020\,(1.7\sigma)$	$0.0060^{+0.0013}_{-0.0013}\,(0.8\sigma)$
Extra source	45.699	50.6830		0.0031 ± 0.0005	
Extra source	45.304	50.7559		0.0022 ± 0.0003	
Extra source	44.807	50.7716		0.0012 ± 0.0007	
Extra source	44.801	50.7753		0.0006 ± 0.0007	
Extra source	44.782	50.8955		0.1716 ± 0.0009	

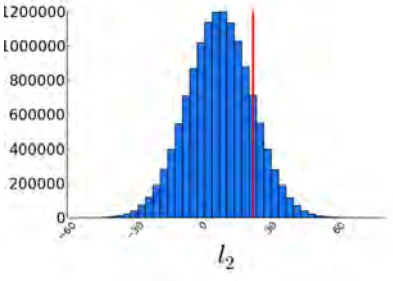
TABLE A.2: Comparison between the CLEAN+source extraction results (shortened to CLN) and the BIRO results for the "QMC" dataset with constant pointing errors.

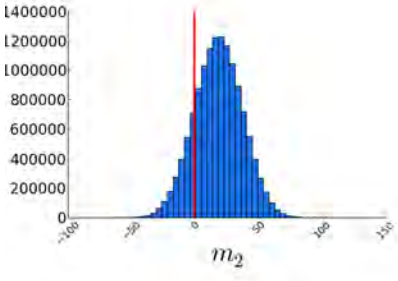
A.3 QMC shape mock data

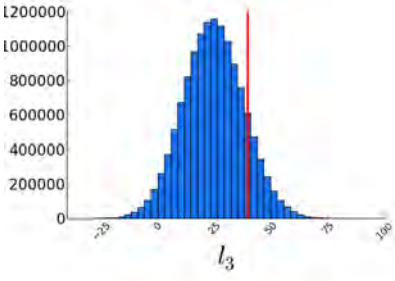
Below are the one-dimensional marginalised posteriors for all the parameters of this simulated dataset, illustrating that the true, input parameters (indicated in each case by a vertical line) are well recovered in all cases. For the pointing errors, l_i indicates the pointing error in right ascension direction for the $i'th$ antenna and similarly, m_i indicates the pointing error in the declination direction for the $i'th$ antenna. The shape parameters for the $j'th$ source are given by $l\text{proj}_j$, $m\text{proj}_j$ and $r = e\text{min}/e\text{maj}_j$ (see Figure 5.8).

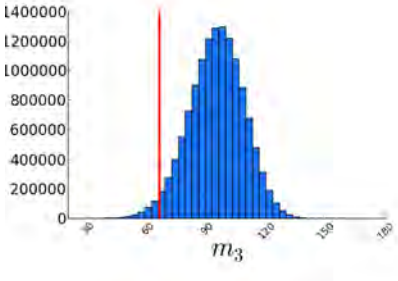

1400000
1200000
1000000
800000
600000
400000
200000
0
l_1

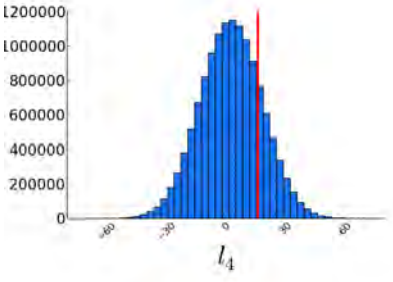

1400000
1200000
1000000
800000
600000
400000
200000
0
m_1

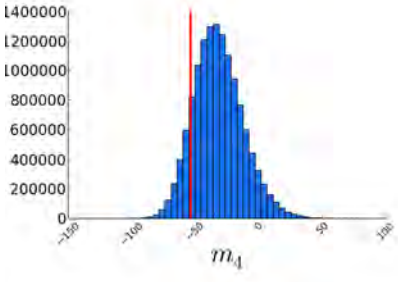

1200000
1000000
800000
600000
400000
200000
0
l_2

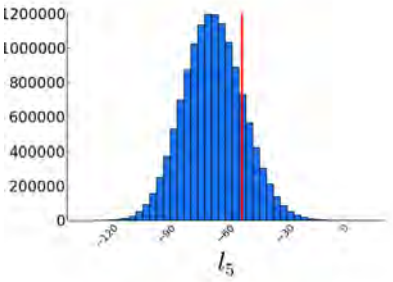

1400000
1200000
1000000
800000
600000
400000
200000
0
m_2

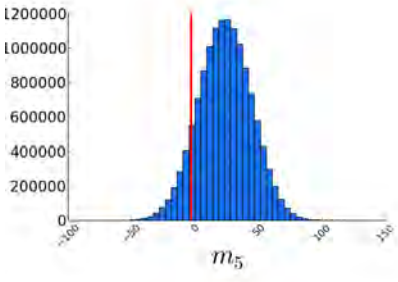

1200000
1000000
800000
600000
400000
200000
0
l_3

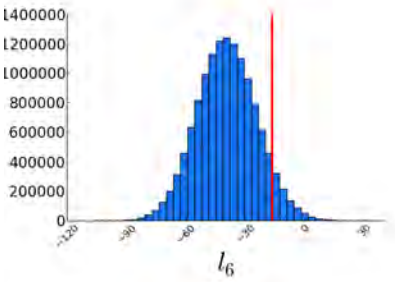

1400000
1200000
1000000
800000
600000
400000
200000
0
m_3

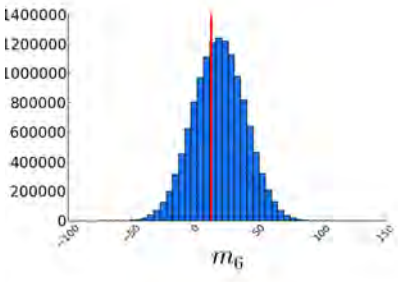

1200000
1000000
800000
600000
400000
200000
0
l_4

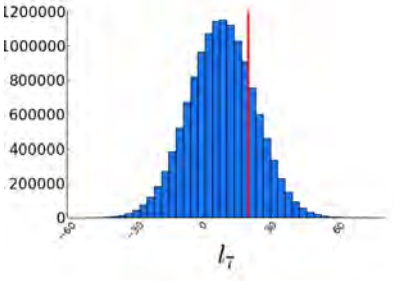

1400000
1200000
1000000
800000
600000
400000
200000
0
m_4

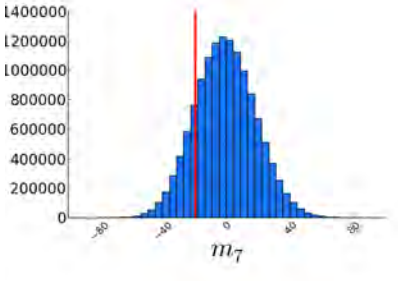

1200000
1000000
800000
600000
400000
200000
0
l_5

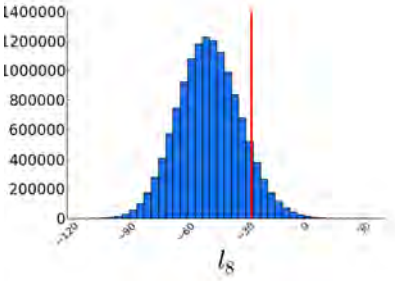

1200000
1000000
800000
600000
400000
200000
0
m_5

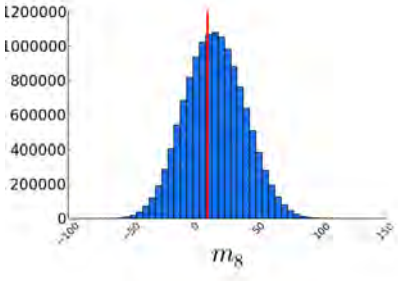

1400000
1200000
1000000
800000
600000
400000
200000
0
l_6

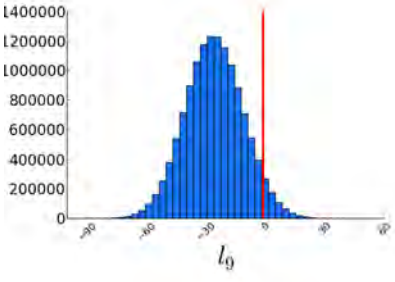

1400000
1200000
1000000
800000
600000
400000
200000
0
m_6

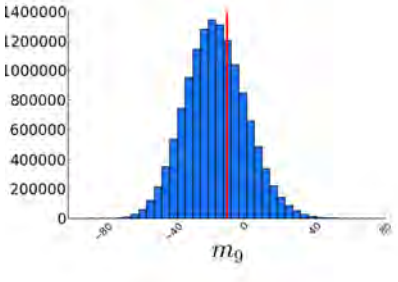

1200000
1000000
800000
600000
400000
200000
0
l_7

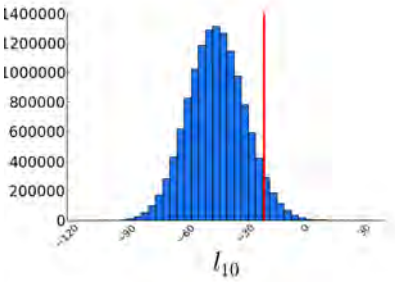

1400000
1200000
1000000
800000
600000
400000
200000
0
m_7

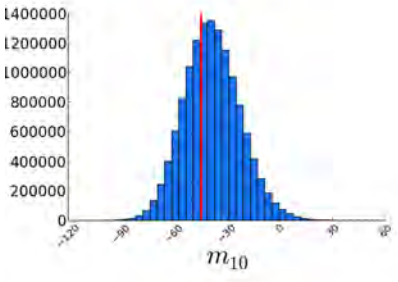

1400000
1200000
1000000
800000
600000
400000
200000
0
l_8

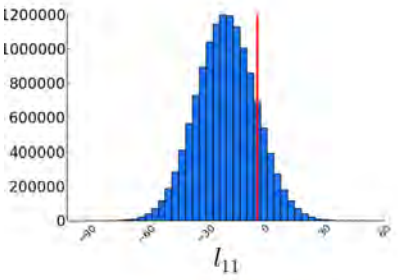

1200000
1000000
800000
600000
400000
200000
0
m_8


1400000
1200000
1000000
800000
600000
400000
200000
0
l_9

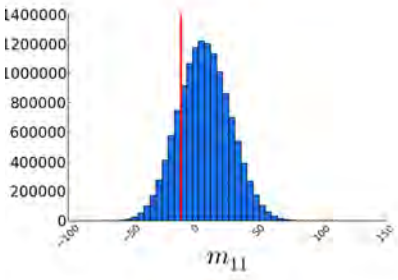

1400000
1200000
1000000
800000
600000
400000
200000
0
m_9


1400000
1200000
1000000
800000
600000
400000
200000
0
l_{10}

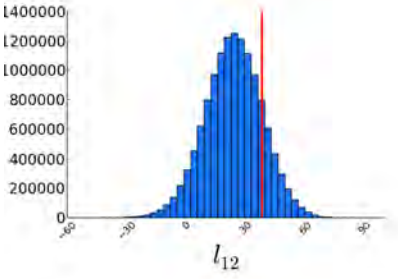

1400000
1200000
1000000
800000
600000
400000
200000
0
m_{10}



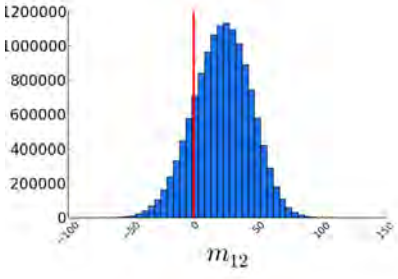
l_{11}



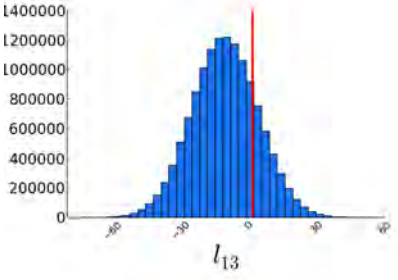
m_{11}



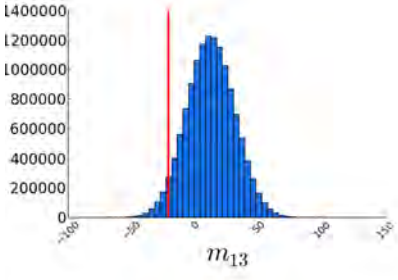
l_{12}



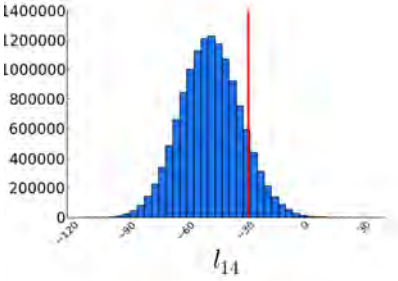
m_{12}



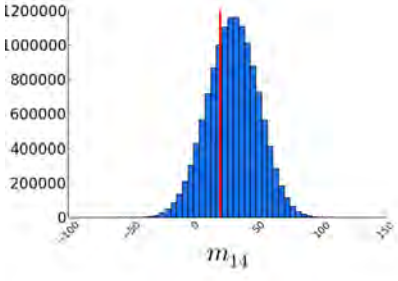
l_{13}



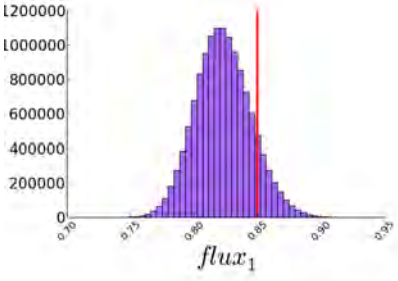
m_{13}



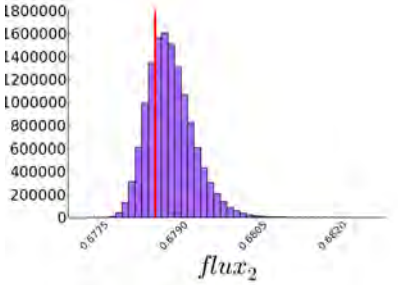
l_{14}



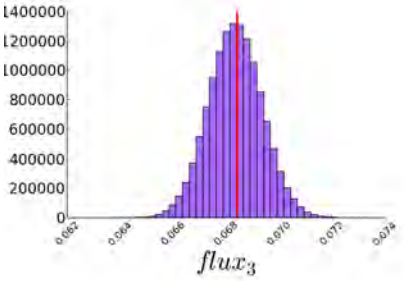
m_{14}

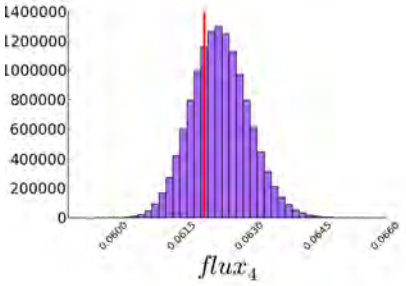


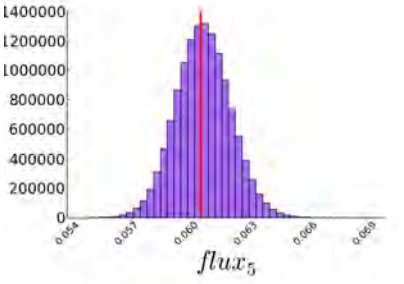
$flux_1$

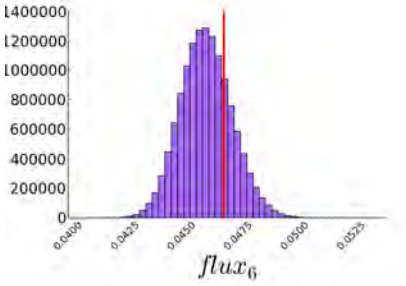


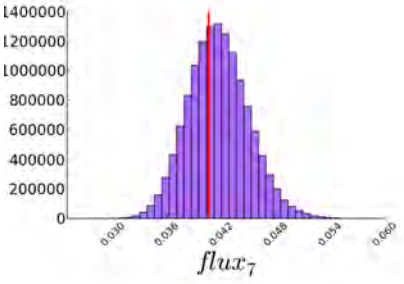
$flux_2$

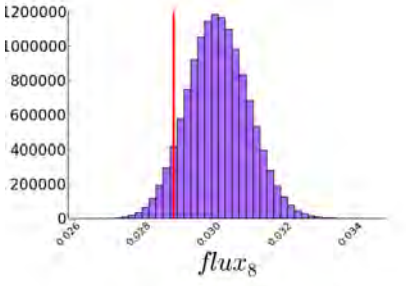

1400000
1200000
1000000
800000
600000
400000
200000
0
0.062
0.064
0.066
0.068
0.070
0.072
0.074
$flux_3$

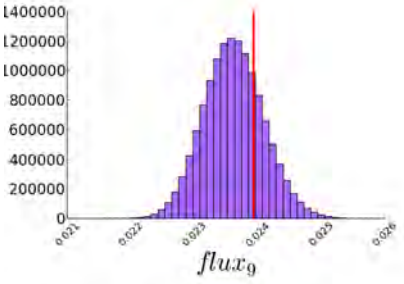

1400000
1200000
1000000
800000
600000
400000
200000
0
0.0600
0.0615
0.0630
0.0645
0.0660
$flux_4$

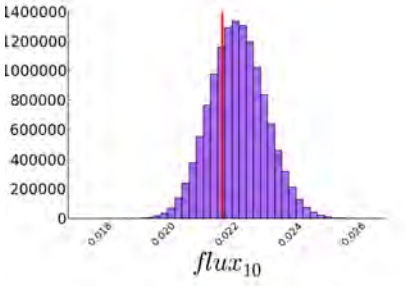

1400000
1200000
1000000
800000
600000
400000
200000
0
0.054
0.057
0.060
0.063
0.066
0.069
$flux_5$

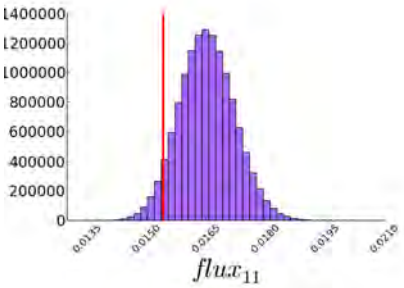

1400000
1200000
1000000
800000
600000
400000
200000
0
0.0400
0.0425
0.0450
0.0475
0.0500
0.0525
$flux_6$

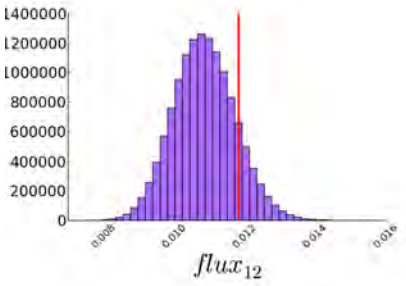

1400000
1200000
1000000
800000
600000
400000
200000
0
0.030
0.036
0.042
0.048
0.054
0.060
$flux_7$

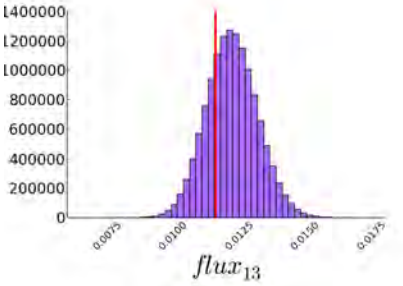

1200000
1000000
800000
600000
400000
200000
0
0.026
0.028
0.030
0.032
0.034
$flux_8$

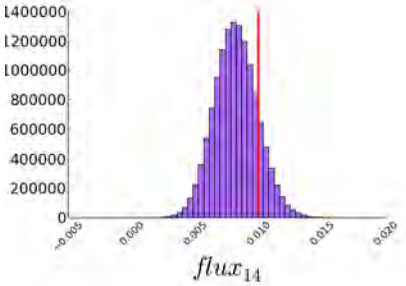

1400000
1200000
1000000
800000
600000
400000
200000
0
0.021
0.022
0.023
0.024
0.025
0.026
$flux_9$

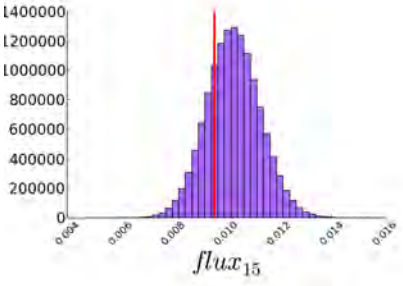

1400000
1200000
1000000
800000
600000
400000
200000
0
0.018
0.020
0.022
0.024
0.026
$flux_{10}$

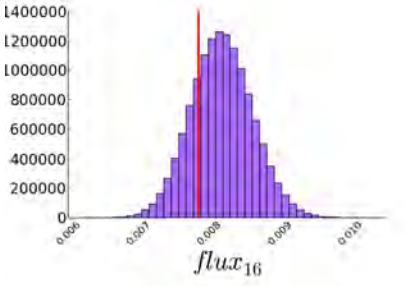

1400000
1200000
1000000
800000
600000
400000
200000
0
0.0135
0.0150
0.0165
0.0180
0.0195
0.0210
$flux_{11}$

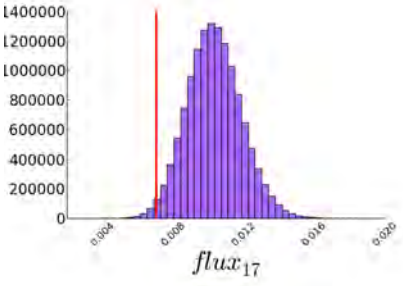

1400000
1200000
1000000
800000
600000
400000
200000
0
0.008
0.010
0.012
0.014
0.016
$flux_{12}$

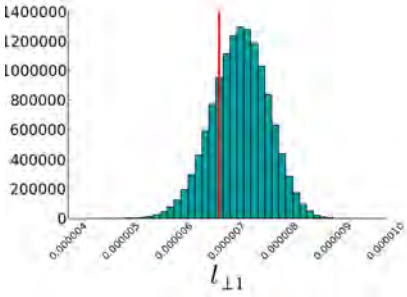

1400000
1200000
1000000
800000
600000
400000
200000
0
0.0075
0.0100
0.0125
0.0150
0.0175
$flux_{13}$

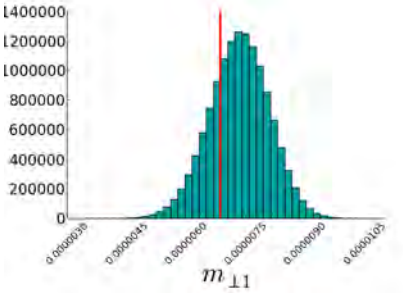

1400000
1200000
1000000
800000
600000
400000
200000
0
-0.005
0.000
0.005
0.010
0.015
0.020
$flux_{14}$

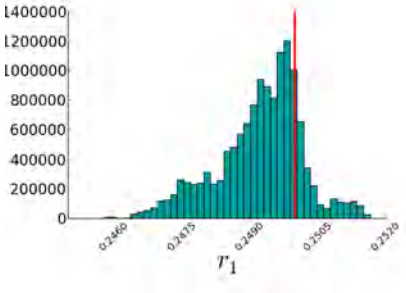

1400000
1200000
1000000
800000
600000
400000
200000
0
0.004
0.006
0.008
0.010
0.012
0.014
0.016
$flux_{15}$

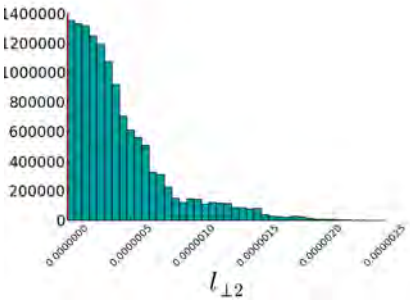

1400000
1200000
1000000
800000
600000
400000
200000
0
0.006
0.007
0.008
0.009
0.010
$flux_{16}$

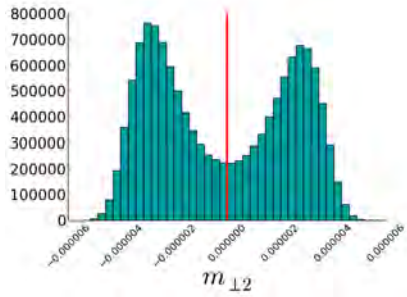

1400000
1200000
1000000
800000
600000
400000
200000
0
0.004
0.008
0.012
0.016
0.020
$flux_{17}$

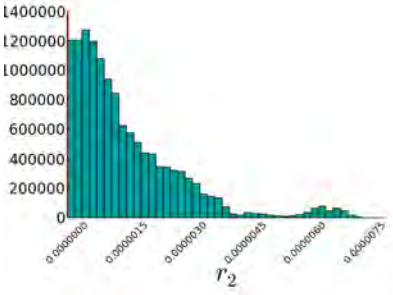

1400000
1200000
1000000
800000
600000
400000
200000
0
0.000004
0.000005
0.000006
0.000007
0.000008
0.000009
0.000010
$l_{\perp 1}$

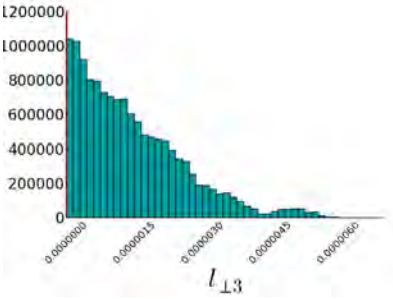

1400000
1200000
1000000
800000
600000
400000
200000
0
0.0000030
0.0000045
0.0000060
0.0000075
0.0000090
0.0000105
$m_{\perp 1}$

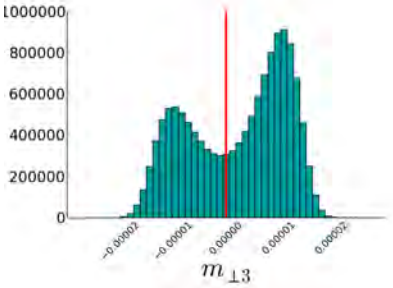

1400000
1200000
1000000
800000
600000
400000
200000
0
0.2460
0.2475
0.2490
0.2505
0.2520
r_1

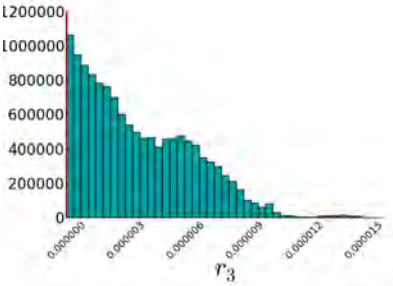

1400000
1200000
1000000
800000
600000
400000
200000
0
0.0000000
0.0000005
0.0000010
0.0000015
0.0000020
0.0000025
$l_{\perp 2}$

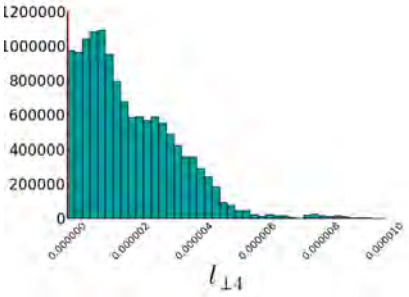

800000
700000
600000
500000
400000
300000
200000
100000
0
-0.000006
-0.000004
-0.000002
0.000000
0.000002
0.000004
0.000006
$m_{\perp 2}$

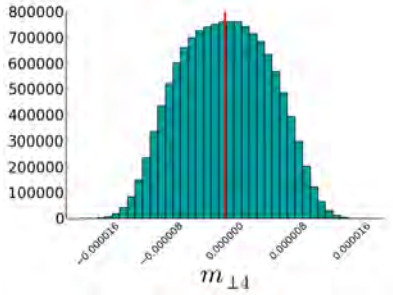

r_2

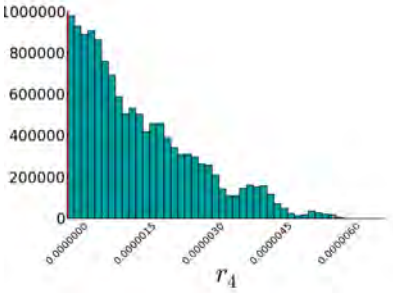

$l_{\perp 3}$

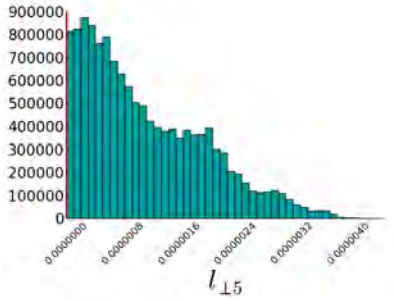

$m_{\perp 3}$

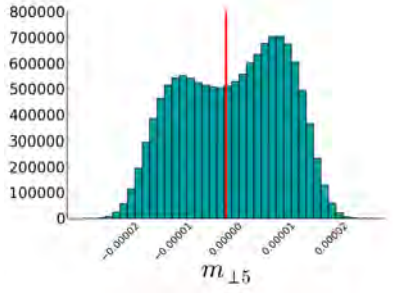

r_3

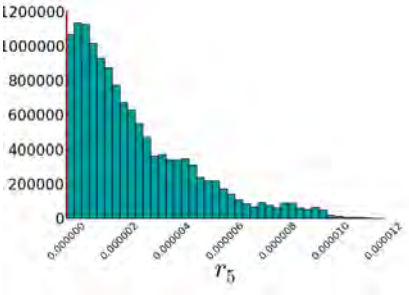

$l_{\perp 4}$

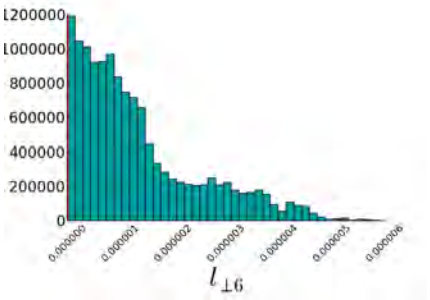

$m_{\perp 4}$

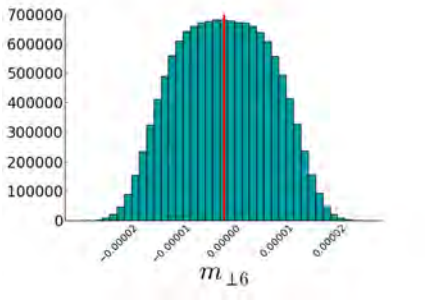

r_4

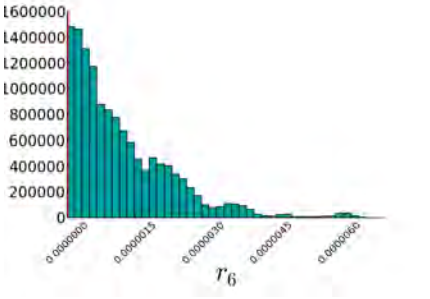

$l_{\perp 5}$

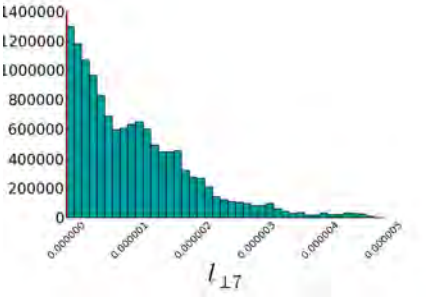

$m_{\perp 5}$

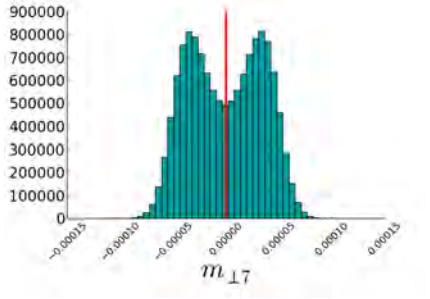

r_5

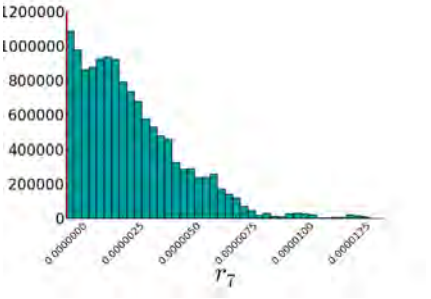

$l_{\perp 6}$

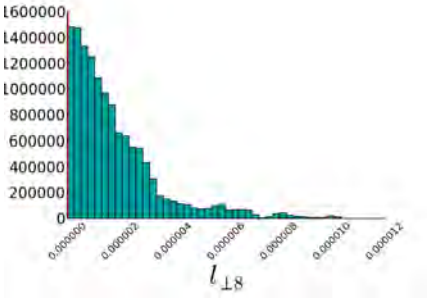

$m_{\perp 6}$

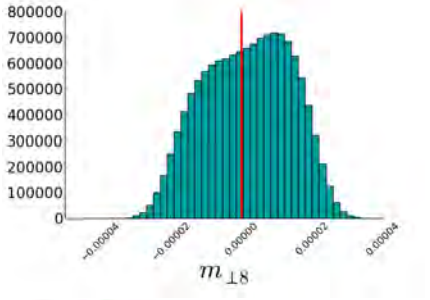

r_6

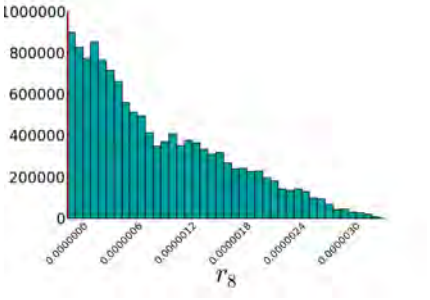

$l_{\perp 7}$

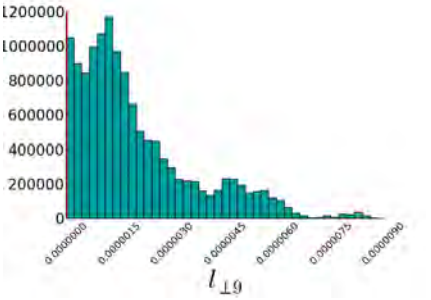

$m_{\perp 7}$

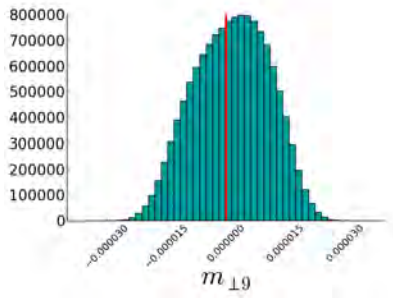

r_7

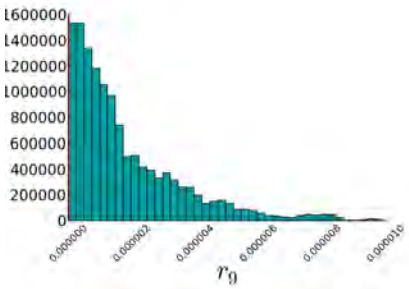

$l_{\perp 8}$

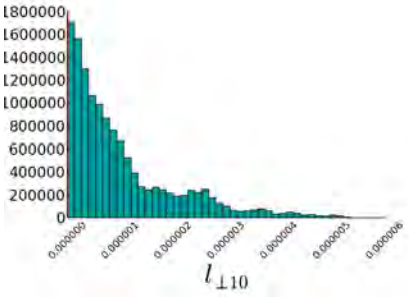

$m_{\perp 8}$

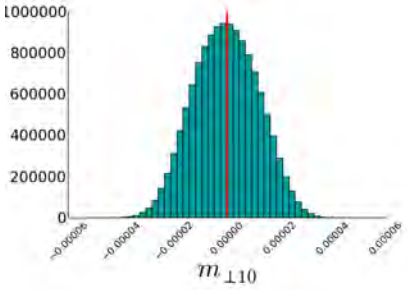

r_8

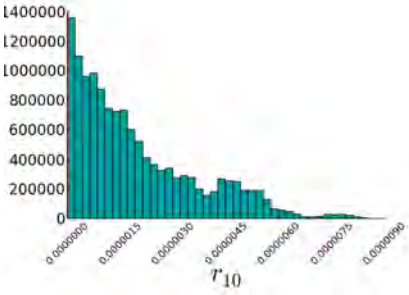

$l_{\perp 9}$

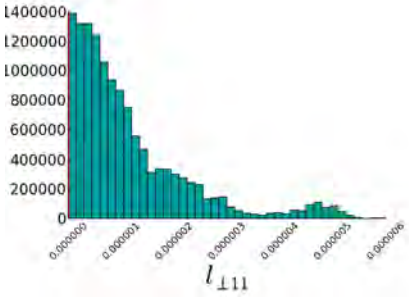

$m_{\perp 9}$

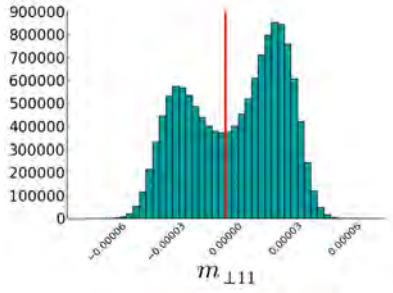

r_9

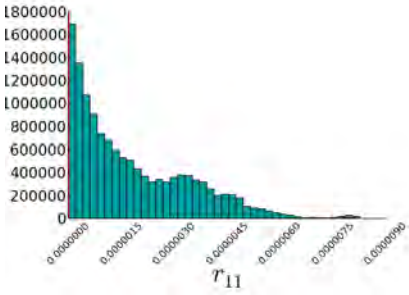

$l_{\perp 10}$

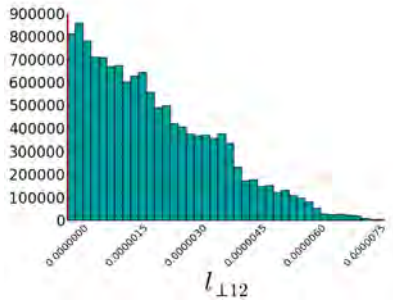

$m_{\perp 10}$

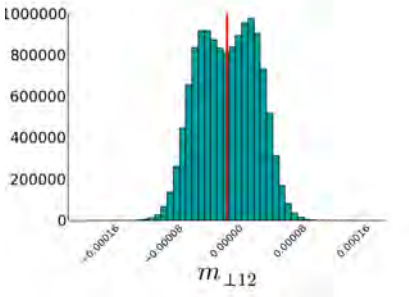

r_{10}

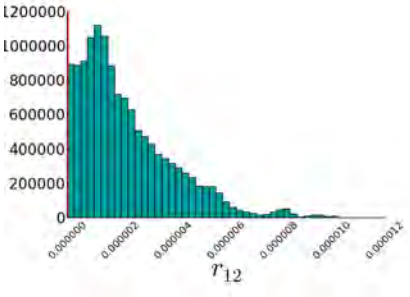

$l_{\perp 11}$

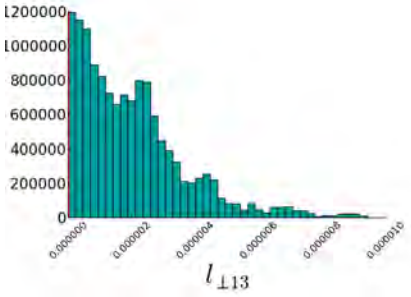

$m_{\perp 11}$

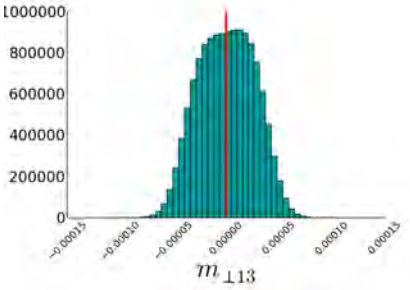

r_{11}

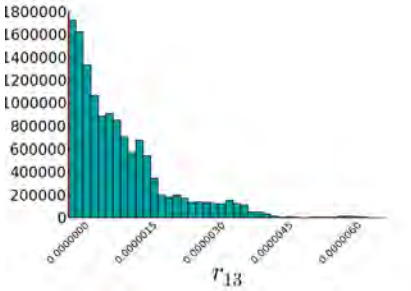

$l_{\perp 12}$

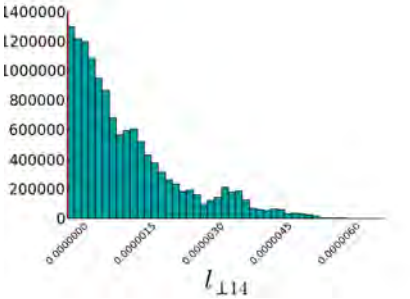

$m_{\perp 12}$

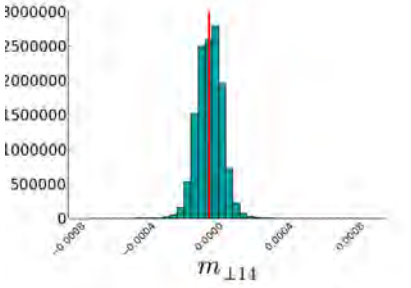

r_{12}

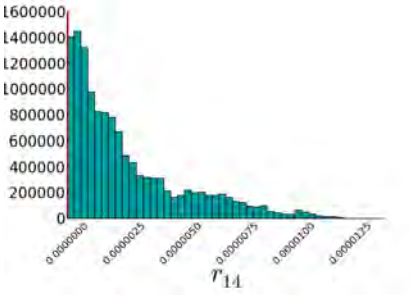

$l_{\perp 13}$

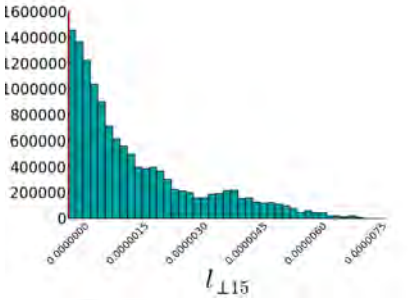

$m_{\perp 13}$

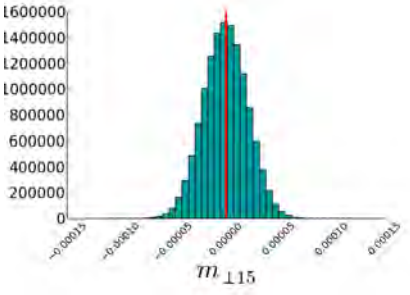

r_{13}

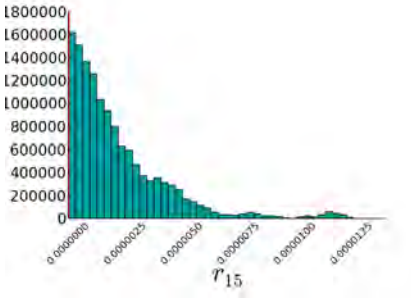

$l_{\perp 14}$

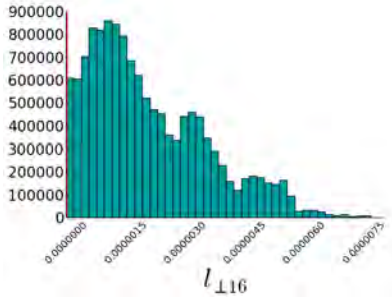

$m_{\perp 14}$

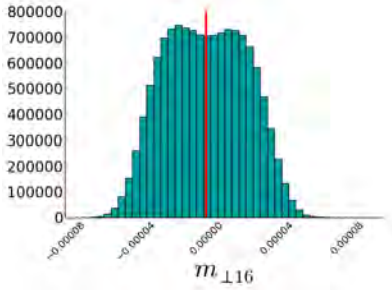

r_{14}

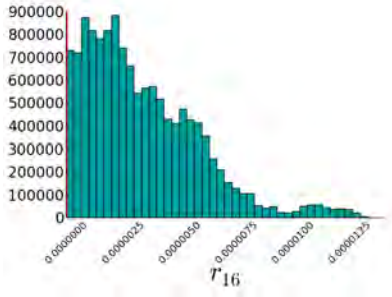

$l_{\perp 15}$

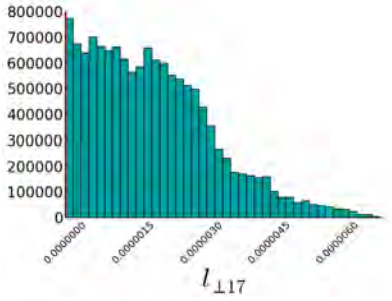

$m_{\perp 15}$

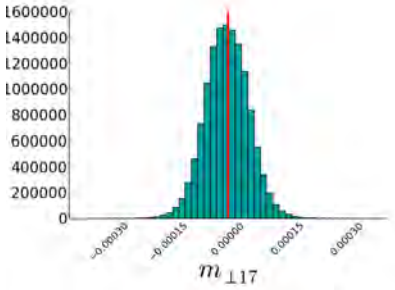

r_{15}

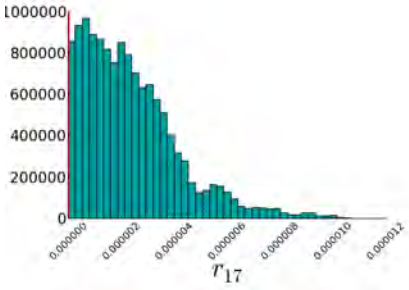

$l_{\perp 16}$

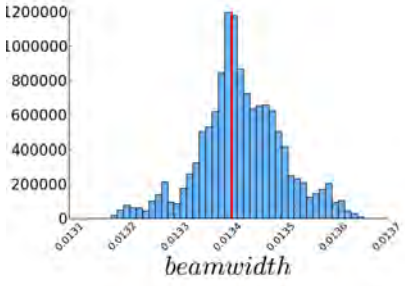

$m_{\perp 16}$

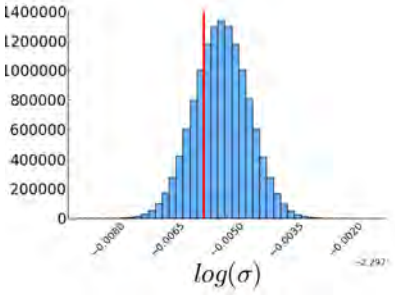

r_{16}


$l_{\perp 17}$


$m_{\perp 17}$


r_{17}


beamwidth


$log(\sigma)$

Name	RA (°)	Dec (°)	True Flux Density (Jy)	CLN Flux Density (Jy)	BIRO Flux Density (Jy)
B292	44.782	50.894	0.8500	$0.6837 \pm 0.0009\,(189.7\sigma)$	$0.8228^{+0.0227}_{-0.0182}\,(1.3\sigma)$
A240	45.303	50.752	0.6786	$0.6763 \pm 0.0003\,(7.6\sigma)$	$0.6788^{+0.0004}_{-0.0004}\,(0.5\sigma)$
C051	45.574	50.904	0.0684	$0.0677 \pm 0.0005\,(1.5\sigma)$	$0.0683^{+0.0010}_{-0.0008}\,(0.2\sigma)$
D221	45.150	50.622	0.0620	$0.0626 \pm 0.0004\,(1.7\sigma)$	$0.0623^{+0.0006}_{-0.0006}\,(0.5\sigma)$
E041	45.655	50.974	0.0607	$0.0601 \pm 0.0007\,(1.0\sigma)$	$0.0609^{+0.0014}_{-0.0014}\,(0.2\sigma)$
F321	45.035	51.019	0.0469	$0.0454 \pm 0.0007\,(2.1\sigma)$	$0.0461^{+0.0011}_{-0.0011}\,(0.7\sigma)$
G223	44.740	50.358	0.0406	$0.0383 \pm 0.0035\,(0.7\sigma)$	$0.0416^{+0.0033}_{-0.0033}\,(0.3\sigma)$
H182	45.339	50.432	0.0290	$0.0291 \pm 0.0007\,(0.2\sigma)$	$0.0302^{+0.0008}_{-0.0008}\,(1.5\sigma)$
I281	44.953	50.839	0.0239	$0.0237 \pm 0.0005\,(0.5\sigma)$	$0.0236^{+0.0004}_{-0.0004}\,(0.8\sigma)$
J091	45.885	50.760	0.0219	$0.0205 \pm 0.0009\,(1.4\sigma)$	$0.0223^{+0.0009}_{-0.0007}\,(0.5\sigma)$
K151	45.662	50.510	0.0154	$0.0164 \pm 0.0008\,(1.4\sigma)$	$0.0165^{+0.0007}_{-0.0007}\,(1.6\sigma)$
L042	45.706	51.058	0.0118	$0.0116 \pm 0.0012\,(0.2\sigma)$	$0.0108^{+0.0008}_{-0.0008}\,(1.3\sigma)$
M072	45.919	50.897	0.0116	$0.0133 \pm 0.0013\,(1.3\sigma)$	$0.0122^{+0.0009}_{-0.0009}\,(0.6\sigma)$
N042	45.851	51.092	0.0100	$0.0128 \pm 0.0023\,(1.2\sigma)$	$0.0080^{+0.0017}_{-0.0013}\,(1.3\sigma)$
O322	44.973	51.139	0.0096	$0.0152 \pm 0.0014\,(3.9\sigma)$	$0.0103^{+0.0010}_{-0.0010}\,(0.8\sigma)$
P111	45.744	50.719	0.0078	$0.0077 \pm 0.0006\,(0.3\sigma)$	$0.0081^{+0.0004}_{-0.0004}\,(0.7\sigma)$
Q182	45.434	50.288	0.0070	$0.0037 \pm 0.0020\,(1.7\sigma)$	$0.0102^{+0.0015}_{-0.0011}\,(2.3\sigma)$
Extra source	45.699	50.6830		0.0031 ± 0.0005	
Extra source	45.304	50.7559		0.0022 ± 0.0003	
Extra source	44.807	50.7716		0.0012 ± 0.0007	
Extra source	44.801	50.7753		0.0006 ± 0.0007	
Extra source	44.782	50.8955		0.1716 ± 0.0009	

TABLE A.3: Comparison between the CLEAN+source extraction results (shortened to CLN) and the BIRO results for the "QMC shape" dataset. The CLEAN+SE results are the same as Table A.2 since the dataset has not changed, only the parameters.

Bibliography

[1] Michelle Lochner (née Knights), et al. Extending BEAMS to incorporate correlated systematic uncertainties. *Journal of Cosmology and Astroparticle Physics*, 2013(01):039, 2013, arXiv:1205.3493.

[2] A. Liddle. *An Introduction to Modern Cosmology*. John Wiley & Sons Ltd, 2003.

[3] S. Dodelson. *Modern Cosmology*. Academic Press, United States of America, 2003.

[4] P. Machamer. *The Cambridge Companion to Galileo*. Cambridge Collections Online. Cambridge University Press, 1998.

[5] J. Cervantes-Cota and G. Smoot. Cosmology todayA brief review. *AIP Conf. Proc.*, 1396:28–52, 2010, arXiv:1107.1789.

[6] V. M. Slipher. Nebulae. *Proceedings of the American Philosophical Society*, 56:403–409, 1917.

[7] E. Hubble. A Relation Between Distance and Radial Velocity Among Extra-Galactic Nebulae. *Proceedings of the National Academy of Sciences*, 15(3), 1929.

[8] J.A. Peacock. *Cosmological Physics*. Cambridge Astrophysics. Cambridge University Press, 1999.

[9] S. Carroll. *An Introduction to General Relativity: Spacetime and Geometry*. Addison Wesley (Pearson Education Inc.), 2004.

[10] S. Weinberg. *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity*. John Wiley & Sons, Inc., 1972.

[11] C. M. Will. The Confrontation between General Relativity and Experiment. *ArXiv e-prints*, 2014, arXiv:1403.7377.

[12] J.A. Wheeler. *Geons, Black Holes, and Quantum Foam: A Life in Physics*. New York: W.W. Norton & Co, 1998.

[13] A Bayesian approach to discrete object detection in astronomical datasets. *Monthly Notices of the Royal Astronomical Society*, 338(3):765–784, 2002, arXiv:0204457.

[14] P. Hobson, G.P. Efstathiou, and A.N. Lasenby. *General Relativity: An Introduction for Physicists*. Cambridge University Press, 2006.

[15] J. D. Barrow, et al. LETTER TO THE EDITOR: On the stability of the Einstein static universe. *Classical and Quantum Gravity*, 20:L155–L164, 2003, arXiv:0302094.

[16] C. Clarkson and R. Maartens. Inhomogeneity and the foundations of concordance cosmology. *Classical and Quantum Gravity*, 27(12):124008, 2010, arXiv:1005.2165.

[17] P. Schneider, J. Ehlers, and E.E. Falco. *Gravitational Lenses*. Astronomy and Astrophysics Library. Springer, 1999.

[18] G. Bernstein. Metric Tests for Curvature from Weak Lensing and Baryon Acoustic Oscillations. *The Astrophysical Journal*, 637:598–607, 2006, arXiv:astro-ph/0503276.

[19] PAR Ade, et al. Planck 2013 results. XVI. Cosmological parameters. 2013, arXiv:arXiv:1303.5076v1.

[20] B. Bassett and R. Hlozek. Baryon acoustic oscillations. In P. Ruiz-Lapuente, editor, *Dark Energy: Observational and Theoretical Approaches*, page 246. Cambridge University Press, 2010, arXiv:0910.5224.

[21] Linda Siobhan Sparke and John Sill Gallagher. *Galaxies in the Universe: An Introduction*. Cambridge University Press, 2007.

[22] A. Conley et al. Supernova Constraints and Systematic Uncertainties from the First Three Years of the Supernova Legacy Survey. *The Astrophysical Journal Supplement Series*, 192(1):1, 2011, arXiv:arXiv:1104.1443.

[23] M. Rowan-Robinson. *Cosmology*. Oxford Physics Series. Clarendon Press, 1977.

[24] J. Roth and J. Primack. Status of cosmological parameters. *Sky & Telescope*, 91(1):20, 1996.

[25] K. Freese. Review of Observational Evidence for Dark Matter in the Universe and in upcoming searches for Dark Stars. *European Astronomical Society Publications Series*, 36:113–126, 2009, arXiv:0812.4005.

[26] A. G. Riess et al. Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant. *The Astronomical Journal*, 116(3):1009, 1998, arXiv:9805201.

[27] S. Perlmutter et al. Measurements of Ω and Λ from 42 High-Redshift Supernovae. *The Astrophysical Journal*, 517(2):565, 1999, arXiv:9812133.

[28] T. P. Sotiriou and V. Faraoni. f(R) theories of gravity. *Reviews of Modern Physics*, 82:451–497, 2010, arXiv:0805.1726.

[29] S. Capozziello and G. Lambiase. A comprehensive view of cosmological Dark Side. *ArXiv e-prints*, 2013, arXiv:1304.5640.

[30] P.C. Davies and S.D. Unwin. Quantum Vacuum Energy And The Masslessness Of The Photon. *Physics Letters B*, 98(274), 1981.

[31] S. Weinberg. Anthropic Bound on the Cosmological Constant. *Phys. Rev. Lett.*, 59:2607–2610, 1987.

[32] E. J. Copeland, M. Sami, and S. Tsujikawa. Dynamics Of Dark Energy. *International Journal of Modern Physics D*, 15(11):1753–1935, 2006, arXiv:0603057.

[33] Carroll, S.M. The Cosmological Constant. *Living Rev. Relativity*, 4, 2001, arXiv:0004075.

[34] Ofer Lahav. Dark energy: how the paradigm shifted. http://physicsworld.com/cws/article/print/2010/jun/02/dark-energy-how-the-paradigm-shifted, 2010.

[35] D. Baumann. TASI Lectures on Inflation. http://arxiv.org/abs/0907.5424, 2009.

[36] A. H. Guth. Inflationary universe: A possible solution to the horizon and flatness problems. *Physical Review D*, 23:347–356, 1981.

[37] Ariel Goobar and Bruno Leibundgut. Supernova Cosmology: Legacy and Future. *Annual Review of Nuclear and Particle Science*, 61(1):251–279, 2011, arXiv:1102.1431.

[38] G. Raffelt. Neutrinos and the stars. *ArXiv e-prints*, 2012, arXiv:1201.1637.

[39] B.W. Carroll and D.A. Ostlie. *An introduction to modern astrophysics*. Pearson Addison-Wesley, 2007.

[40] D. Moore. Supernovae. http://www.talkorigins.org/faqs/supernova/, 2003.

[41] S. Perlmutter, et al. Cosmology From Type IA Supernovae: Measurements, Calibration Techniques, and Implications. In *American Astronomical Society Meeting Abstracts*, volume 29 of *Bulletin of the American Astronomical Society*, page 1351, 1997, arXiv:9812473.

[42] R. Amanullah et al. Spectra and Hubble Space Telescope Light Curves of Six Type Ia Supernovae at $0.511 < z < 1.12$ and the Union2 Compilation. *The Astrophysical Journal*, 716(1):712, 2010, arXiv:1004.1711.

[43] E. Komatsu, et al. Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation. *The Astrophysical Journal Supplement Series*, 192(2):18, 2011, arXiv:1001.4538.

[44] Will J. Percival, et al. Baryon acoustic oscillations in the Sloan Digital Sky Survey Data Release 7 galaxy sample. *Monthly Notices of the Royal Astronomical Society*, 401(4):2148–2168, 2010, arXiv:0907.1660.

[45] M. Kunz, B. Bassett and R. Hlozek. Bayesian Estimation Applied to Multiple Species. *Phys.Rev.D*, 75:103508, 2007, arXiv:0611004v2.

[46] R. Hlozek et al. Photometric Supernova Cosmology with BEAMS and SDSS-II. *The Astrophysical Journal*, 752:79, 2012, arXiv:1111.5328.

[47] J. Newling et al. Parameter estimation with Bayesian estimation applied to multiple species in the presence of biases and correlations. *Monthly Notices of the Royal Astronomical Society*, 421(2):913–925, 2012, arXiv:1110.6178.

[48] Saleem Zaroubi. The epoch of reionization. In Tommy Wiklind, Bahram Mobasher, and Volker Bromm, editors, *The First Galaxies*, volume 396 of *Astrophysics and Space Science Library*, pages 45–101. Springer Berlin Heidelberg, 2013, arXiv:1206.0267v1.

[49] J E Gunn and B A Peterson. On the density of neutral hydrogen in intergalactic space. *The Astrophysical Journal*, 142:1633, 1965.

[50] S. Zaroubi. Probing the Epoch of Reionization with Low Frequency Arrays. *ArXiv e-prints*, 2010, arXiv:1002.2667.

[51] M Colless, et al. The 2df galaxy redshift survey: Spectra and redshifts. *Monthly Notices of the Royal Astronomical Society*, 328(4):1039–1063, 2001, arXiv:0106498.

[52] J. Jeans. The Stability of a Spherical Nebula. *Phil. Trans. R. Soc. Lond. A*, 199:1–53, 1902.

[53] S. T. Myers, et al. The Billion Galaxy Cosmological HI Large Deep Survey (BiG-CHILDS). In *astro2010: The Astronomy and Astrophysics Decadal Survey*, volume 2010 of *Astronomy*, page 219, 2009, arXiv:0903.0615.

[54] Stephen D Landy and Alexander S Szalay. Bias and variance of angular correlation functions. *The Astrophysical Journal*, 412:64, 1993.

[55] Robert C. Nichol. Cosmology with galaxy correlations. *General Relativity and Gravitation*, 40(2-3):249–267, 2007, arXiv:0708.2824.

[56] P. Tozzi. Cosmological Parameters from Galaxy Clusters: An Introduction. In L. Papantonopoulos, editor, *The Invisible Universe: Dark Matter and Dark Energy*, volume 720 of *Lecture Notes in Physics, Berlin Springer Verlag*, page 125, 2007, arXiv:astro-ph/0602072.

[57] Daniel J. Eisenstein, et al. Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies. *The Astrophysical Journal*, 633(2):560–574, 2005, arXiv:0501171.

[58] Ariel G. Sánchez, et al. The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological implications of the large-scale two-point correlation function. *Monthly Notices of the Royal Astronomical Society*, 425(1):415–437, 2012, arXiv:1203.6616.

[59] C. L. Carilli and S. Rawlings. Motivation, key science projects, standards and assumptions. *New Astronomy Reviews*, 48:979–984, 2004, arXiv:0409274.

[60] P. Bull, et al. Late-time cosmology with 21cm intensity mapping experiments. *ArXiv e-prints*, 2014, arXiv:1405.1452.

[61] L. Wolz, et al. The effect of foreground subtraction on cosmological measurements from Intensity Mapping. *ArXiv e-prints*, 2013, arXiv:1310.8144.

[62] J.J. Condon and S.M. Ransom. Essential radio astronomy course notes, 2010.

[63] C. M. Urry and P. Padovani. Unified Schemes for Radio-Loud Active Galactic Nuclei. *Publications of the Astronomical Society of the Pacific*, 107:803, 1995, arXiv:9506063.

[64] B. L. Fanaroff and J. M. Riley. The morphology of extragalactic radio sources of high and low luminosity. *Monthly Notices of the Royal Astronomical Society*, 167:31P–36P, 1974.

[65] M. Kramer, et al. Strong-field tests of gravity using pulsars and black holes. *New Astronomy Reviews*, 48:993–1002, 2004, arXiv:astro-ph/0409379.

[66] T.L. Wilson, K. Rohlfs, and S. Hüttemeister. *Tools of Radio Astronomy*. Astronomy and Astrophysics Library. Springer, 2009.

[67] A.R. Thompson, J.M. Moran, and G.W. Swenson. *Interferometry and Synthesis in Radio Astronomy*. A Wiley-Interscience publication. Wiley, 2001.

[68] K.G. Jansky. Electrical Disturbances Apparently of Extraterrestrial Origin. *Proceedings of the Institute of Radio Engineers*, 21, 1933.

[69] O. Smirnov. Interferometry & the radio interferometry measurement equation (rime). 2012.

[70] J. Pety. Imaging, deconvolution and image analysis. 2010.

[71] O. M. Smirnov. Revisiting the radio interferometer measurement equation. I. A full-sky Jones formalism. *Astronomy and Astrophysics*, 527:A106, 2011, arXiv:1101.1764.

[72] J. P. Hamaker, J. D. Bregman, and R. J. Sault. Understanding radio polarimetry. I. Mathematical foundations. *Astronomy and Astrophysics Supplement*, 117:137–147, 1996.

[73] J. P. Hamaker and J. D. Bregman. Understanding radio polarimetry. III. Interpreting the IAU/IEEE definitions of the Stokes parameters. *Astronomy and Astrophysics Supplement*, 117:161–165, 1996.

[74] R. C. Jones. New calculus for the treatment of optical systems. I. Description and discussion of the calculus. *Journal of the Optical Society of America (1917-1983)*, 31:488, 1941.

[75] T Bayes. An essay towards solving a problem in the doctrine of chances. *Phil. Trans. Roy. Soc.*, 53:370418, (Reproduced in: Biometrika, 45, 293-315 (1958)) (1763).

[76] R. Trotta. Bayes in the sky: Bayesian inference and model selection in cosmology. *Contemporary Physics*, 49:71–104, 2008, arXiv:0803.4089.

[77] E.T. Jaynes. *Probability Theory. The logic of science.* Cambridge University Press, 2003.

[78] D. MacKay. *Information Theory, Inference, and Learning Algorithms.* Cambridge University Press, 2003.

[79] N. Metropolis et al. Equation of State Calculations by Fast Computing Machines. *J. Chem. Phys.*, 21:1087, 1953.

[80] W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.

[81] W.R. Gilks, S. Richardson, and D. Spiegelhalter. *Markov Chain Monte Carlo in Practice.* Chapman & Hall/CRC Interdisciplinary Statistics. Taylor & Francis, 1995.

[82] A. Heavens. Statistical techniques in cosmology. *ArXiv e-prints*, 2009, arXiv:0906.0664.

[83] M. Cowles and B. Carlin. Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review. *Journal of the American Statistical Association*, 91(434):883–904, 1994.

[84] A. Gelman and D.B Rubin. Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, 7:457–472, 1992.

[85] William H. Press, et al. *Numerical Recipes in C: The Art of Scientific Computing (second edition).* Cambridge University Press, 1992.

[86] R.A. Fisher. The Logic of Inductive Inference. *J. Roy. Stat. Soc.*, 98, 1935.

[87] B. Bassett et al. Fisher Matrix Preloaded - Fisher4Cast. *International Journal of Modern Physics D*, 20(13):2559–2598, 2011, arXiv:0906.0993.

[88] T. Matsubara. Correlation Function in Deep Redshift Space as a Cosmological Probe. *The Astrophysical Journal*, 615(2):573, 2004, arXiv:0408349.

[89] A. Leon-Garcia. *Probability, Statistics and Random Processes for Electrical Engineering.* Prentice Hall, 2007.

[90] M. Chevallier and D. Polarski. Accelerating Universes With Scaling Dark Matter. *International Journal of Modern Physics D*, 10(02):213–223, 2001, arXiv:0009008.

[91] E. Linder. Exploring the Expansion History of the Universe. *Phys. Rev. Lett.*, 90:091301, 2003, arXiv:0208512.

[92] The Dark Energy Survey Collaboration. The Dark Energy Survey. *ArXiv e-prints*, 2005, arXiv:0510346.

[93] N. Kaiser and the Pan-STARRS Team. The Pan-STARRS Survey Telescope Project. *Bulletin of the American Astronomical Society*, 37:1409, 2005.

[94] LSST Science Collaboration. LSST Science Book, Version 2.0. *ArXiv e-prints*, 2009, arXiv:0912.0201.

[95] B. L. Falck, A. G. Riess, and R. Hlozek. Characterizing the Contaminating Distance Distribution for Bayesian Supernova Cosmology. *The Astrophysical Journal*, 723:398–408, 2010, arXiv:1009.1903.

[96] S. Jha, A.G. Riess, and R.P. Kirshner. Improved Distances to Type Ia Supernovae with Multicolor Light-Curve Shapes: MLCS2k2. *The Astrophysical Journal*, 659:122–148, 2007, arXiv:0612666.

[97] J. Guy et al. SALT2: using distant supernovae to improve the use of type Ia supernovae as distance indicators. *Astronomy and Astrophysics*, 466:11–21, 2007, arXiv:0701828.

[98] M. Sako et al. Photometric Type Ia Supernova Candidates from the Three-year SDSS-II SN Survey Data. *The Astrophysical Journal*, 738(2):162, 2011, arXiv:1107.5106.

[99] D. Poznanski, et al. Not Color-Blind: Using Multiband Photometry to Classify Supernovae. *The Publications of the Astronomical Society of the Pacific*, 114:833–845, 2002, arXiv:astro-ph/0202198.

[100] M. Sako et al. The Sloan Digital Sky Survey-II Supernova Survey: Search Algorithm and Follow-up Observations. *The Astronomical Journal*, 135(1):348, 2008, arXiv:0708.2750.

[101] E. Gjergo, et al. Type Ia supernovae selection and forecast of cosmology constraints for the Dark Energy Survey. *Astroparticle Physics*, 42:52–61, 2013, arXiv:1205.1480.

[102] J. P. Bernstein et al. Supernova Simulations and Strategies for the Dark Energy Survey. *The Astrophysical Journal*, 753(2):152, 2012, arXiv:1111.1969.

[103] J. A. Frieman et al. The Sloan Digital Sky Survey-II Supernova Survey: Technical Summary. *The Astronomical Journal*, 135(1):338, 2008.

[104] R. Kessler et al. SNANA: A Public Software Package for Supernova Analysis. *Publications of the Astronomical Society of the Pacific*, 121:1028–1035, 2009, arXiv:0908.4280.

[105] R. Kessler et al. First-Year Sloan Digital Sky Survey-II Supernova Results: Hubble Diagram and Cosmological Parameters. *The Astrophysical Journal Supplement Series*, 185(1):32, 2009, arXiv:0908.4274.

[106] L. Hui and P. Greene. Correlated fluctuations in luminosity distance and the importance of peculiar motion in supernova surveys. *Phys. Rev. D*, 73:123526, 2006, arXiv:0512159.

[107] A. Cooray and R.R. Caldwell. Large-scale bulk motions complicate the Hubble diagram. *Phys. Rev. D*, 73:103002, 2006, arXiv:0601377.

[108] C. Gordon, K. Land, and A. Slosar. Cosmological Constraints from Type Ia Supernovae Peculiar Velocity Measurements. *Phys. Rev. Lett.*, 99:081301, 2007, arXiv:0705.1718.

[109] T.M. Davis et al. The Effect of Peculiar Velocities on Supernova Cosmology. *The Astrophysical Journal*, 741(1):67, 2011, arXiv:1012.2912.

[110] N. Palanque-Delabrouille et al. Photometric redshifts for type Ia supernovae in the supernova legacy survey. *Astronomy and Astrophysics*, 514:A63, 2010, arXiv:0911.1629.

[111] M. Doi et al. Photometric Response Functions of the Sloan Digital Sky Survey Imager. *The Astronomical Journal*, 139:1628–1648, 2010, arXiv:1002.3701.

[112] U. Feindt, M. Kowalski, and K. Paech. The self-calibrating Hubble diagram. *Journal of Cosmology and Astroparticle Physics*, 2012(04):001, 2012, arXiv:1201.0765.

[113] R. J. Foley et al. A Mismatch in the Ultraviolet Spectra between Low-redshift and Intermediate-redshift Type Ia Supernovae as a Possible Systematic Uncertainty for Supernova Cosmology. *The Astronomical Journal*, 143(5):113, 2012, arXiv:1010.2749.

[114] K. S. Mandel, G. Narayan, and R. P. Kirshner. Type Ia Supernova Light Curve Inference: Hierarchical Models in the Optical and Near-infrared. *The Astrophysical Journal*, 731(2):120, 2011, arXiv:1011.5910.

[115] C. Gunnarsson et al. Corrections for Gravitational Lensing of Supernovae: Better than Average? *The Astrophysical Journal*, 640(1):417, 2006, arXiv:0506764.

[116] J. Jönsson et al. Weighing dark matter haloes with gravitationally lensed supernovae. *Monthly Notices of the Royal Astronomical Society*, 402(1):526–536, 2010, arXiv:0910.4098.

[117] L. Amendola et al. Large-Scale Inhomogeneities May Improve the Cosmic Concordance of Supernovae. *Phys. Rev. Lett.*, 105:121302, 2010, arXiv:1002.1232.

[118] P. Zhang and P. S. Corasaniti. Cosmic Dust Induced Flux Fluctuations: Bad and Good Aspects. *The Astrophysical Journal*, 657(1):71, 2007, arXiv:0607635.

[119] P. S. Corasaniti. The impact of cosmic dust on supernova cosmology. *Monthly Notices of the Royal Astronomical Society*, 372:191198, 2006, arXiv:0603833.

[120] P. L. Kelly et al. Hubble Residuals of Nearby Type Ia Supernovae are Correlated with Host Galaxy Masses. *The Astrophysical Journal*, 715(2):743, 2010, arXiv:0912.0929.

[121] R. R. Gupta et al. Improved Constraints on Type Ia Supernova Host Galaxy Properties Using Multi-wavelength Photometry and Their Correlations with Supernova Properties. *The Astrophysical Journal*, 740(2):92, 2011, arXiv:1107.6003.

[122] C. B. D'Andrea et al. Spectroscopic Properties of Star-forming Host Galaxies and Type Ia Supernova Hubble Residuals in a nearly Unbiased Sample. *The Astrophysical Journal*, 743(2):172, 2011, arXiv:1110.5517.

[123] J. Meyers et al. The Hubble Space Telescope Cluster Supernova Survey. III. Correlated Properties of TypeIa Supernovae and Their Hosts at $0.9 < Z < 1.46$. *The Astrophysical Journal*, 750(1):1, 2012, arXiv:1201.3989.

[124] E. Komatsu et al. Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation. *The Astrophysical Journal Supplement Series*, 192(2):18, 2011, arXiv:1001.4538v3.

[125] A. Kim and E. Linder. Correlated Supernova Systematics and Ground Based Surveys. *JCAP*, 06(020), 2011, arXiv:1102.1992v1.

[126] N. Suzuki et al. The Hubble Space Telescope Cluster Supernova Survey. V. Improving the Dark-energy Constraints above $z > 1$ and Building an Early-type-hosted Supernova Sample. *The Astrophysical Journal*, 746(1):85, 2012, arXiv:1105.3470.

[127] E. Ising. Beitrag zur Theorie des Ferromagnetismus. *Z. Phys.*, 31:253, 1925.

[128] William A. Link and Mitchell J. Eaton. On thinning of chains in MCMC. *Methods in Ecology and Evolution*, 3(1):112–115, 2012.

[129] M.C. March et al. Improved constraints on cosmological parameters from SNIa data. *MNRAS*, 418(4):23082329, 2011, arXiv:1102.3237.

[130] J. A. Högbom. Aperture Synthesis with a Non-Regular Distribution of Interferometer Baselines. *Astronomy and Astrophysics Supplement*, 15:417, 1974.

[131] H. Junklewitz, et al. RESOLVE: A new algorithm for aperture synthesis imaging of extended emission in radio astronomy. *ArXiv e-prints*, 2013, arXiv:1311.5282.

[132] E. C. Sutton and B. D. Wandelt. Optimal Image Reconstruction in Radio Interferometry. *The Astrophysical Journal Supplement Series*, 162:401–416, 2006, arXiv:0604331.

[133] T. J. Cornwell. Multiscale CLEAN Deconvolution of Radio Synthesis Images. *IEEE Journal of Selected Topics in Signal Processing*, 2:793–801, 2008.

[134] J. G. Ables. Maximum Entropy Spectral Analysis. *Astronomy and Astrophysics Supplement*, 15:383, 1974.

[135] S. F. Gull and G. J. Daniell. Image reconstruction from incomplete and noisy data. *Nature*, 272:686–690, 1978.

[136] R. K. Bryan and J. Skilling. Deconvolution by maximum entropy, as illustrated by application to the jet of M87. *Monthly Notices of the Royal Astronomical Society*, 191:69–79, 1980.

[137] H. Junklewitz, M. A. Bell, and T. Enßlin. A new approach to multi-frequency synthesis in radio interferometry. *ArXiv e-prints*, 2014, arXiv:1401.4711.

[138] I. Martí-Vidal, et al. UVMULTIFIT: A versatile tool for fitting astronomical radio interferometric data. *ArXiv e-prints*, 2014, arXiv:1401.4984.

[139] M. Selig and T. Enßlin. D3PO - Denoising, Deconvolving, and Decomposing Photon Observations. *ArXiv e-prints*, 2013, arXiv:1311.1888.

[140] AMI Consortium, et al. AMI SZ observations and Bayesian analysis of a sample of six redshift-one clusters of galaxies. *ArXiv e-prints*, 2012, arXiv:1210.7771.

[141] P. Carvalho, G. Rocha, and M. P. Hobson. A fast Bayesian approach to discrete object detection in astronomical data sets - PowellSnakes I. *Monthly Notices of the Royal Astronomical Society*, 393:681–702, 2009, arXiv:0802.3916.

[142] P. Carvalho, et al. PowellSnakes II: a fast Bayesian approach to discrete object detection in multi-frequency astronomical data sets. *Monthly Notices of the Royal Astronomical Society*, 427:1384–1400, 2012, arXiv:1112.4886.

[143] J. T. L. Zwart, et al. The Arcminute Microkelvin Imager. *Monthly Notices of the Royal Astronomical Society*, 391:1545–1558, 2008, arXiv:0807.2469.

[144] F. Feroz, et al. Bayesian modelling of clusters of galaxies from multifrequency-pointed Sunyaev-Zel'dovich observations. *Monthly Notices of the Royal Astronomical Society*, 398:2049–2060, 2009, arXiv:0811.1199.

[145] J. T. L. Zwart, et al. Sunyaev-Zel'dovich observations of galaxy clusters out to the virial radius with the Arcminute Microkelvin Imager. *Monthly Notices of the Royal Astronomical Society*, 418:2754–2772, 2011, arXiv:1008.0443.

[146] K. Lancaster, et al. Very Small Array observations of the Sunyaev-Zel'dovich effect in nearby galaxy clusters. *Monthly Notices of the Royal Astronomical Society*, 359:16–30, 2005, arXiv:astro-ph/0405582.

[147] M. Born and E. Wolf. *Principles of Optics Electromagnetic Theory of Propagation, Interference and Diffraction of Light 2nd edition*. 1964.

[148] J. E. Noordam and O. M. Smirnov. The MeqTrees software system and its use for third-generation calibration of radio interferometers. *Astronomy and Astrophysics*, 524:A61, 2010, arXiv:1101.1745.

[149] Laura Dietz. Directed factor graph notation for generative models. *Max Planck Institute for Informatics, Tech. Rep*, 2010.

[150] O. M. Smirnov. Revisiting the radio interferometer measurement equation. II. Calibration and direction-dependent effects. *Astronomy and Astrophysics*, 527:A107, 2011, arXiv:1101.1765.

[151] O. M. Smirnov and A. G. de Bruyn. MeqTrees and direction-dependent effects. *ArXiv e-prints*, 2011, arXiv:1110.2916.

[152] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013.

[153] P. Schneider. Weak Gravitational Lensing with Ska. In M. P. van Haarlem, editor, *Perspectives on Radio Astronomy: Science with Large Antenna Arrays*, page 203, 2000, arXiv:9907146.

[154] P. Patel, et al. Weak Lensing Measurements in Simulations of Radio Images. *ArXiv e-prints*, 2013, arXiv:1303.4650.

[155] J. Jasche and B. D. Wandelt. Bayesian inference from photometric redshift surveys. *Monthly Notices of the Royal Astronomical Society*, 425:1042–1056, 2012, arXiv:1106.2757.

[156] S. Duane, et al. Hybrid Monte Carlo. *Physics Letters B*, 195:216–222, 1987.

[157] R. M. Neal. *Handbook of Markov Chain Monte Carlo*, chapter MCMC using Hamiltonian dynamics. Chapman & Hall/CRC, 2012, arXiv:1206.1901.

[158] J. Goodman and J. Weare. Ensemble samplers with affine invariance. *Communications in Applied Mathematics and Computational Science*, 5(1):65–80, 2010.

[159] D. Foreman-Mackey, et al. emcee: The MCMC Hammer. *Publications of the Astronomical Society of the Pacific*, 125:306–312, 2013, arXiv:1202.3665.